不等式约束极大极小问题的可行下降束方法

简金宝^{①②}, 唐春明^{②*}, 唐菲^②

① 玉林师范学院数学与信息科学学院, 广西高校复杂系统优化与大数据处理重点实验室, 玉林 537000;

② 广西大学数学与信息科学学院, 南宁 530004

E-mail: jianjb@gxu.edu.cn, cmtang@gxu.edu.cn, tangfei18@126.com

收稿日期: 2015-01-06; 接受日期: 2015-08-01; * 通信作者

广西自然科学基金 (批准号: 2013GXNSFAA019013 和 2014GXNSFFA118001) 和国家自然科学基金 (批准号: 11271086 和 11301095) 资助项目

摘要 本文提出一个求解不等式约束极大极小问题的可行下降束方法. 该方法的主要特点有 (1) 借助于函数的次梯度及束方法思想, 不需要假设原问题的分量函数具备光滑性; (2) 利用部分割平面模型技术, 每次无效步迭代仅需利用一个分量函数的函数值和次梯度产生新的割平面, 从而有效减少了计算量; (3) 能够保证有效迭代点的可行性及目标函数的下降性; (4) 引入次梯度聚集技术, 对束集中的次梯度进行聚集, 克服了数值计算和存储的困难; (5) 算法具备全局收敛性, 且初步的数值试验表明算法是有效的.

关键词 不等式约束 极大极小问题 可行下降 束方法 部分割平面模型

MSC (2010) 主题分类 90C30, 65K05

1 引言

本文研究求解如下不等式约束极大极小 (Minimax) 问题:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} f(x) \\ \text{s.t. } c(x) \leq 0, \end{aligned} \quad (1.1)$$

其中, 目标函数 $f(x) = \max\{f_i(x), i \in I\}$, $I = \{1, \dots, m\}$, 约束函数 $c(x) = \max\{c_j(x), j \in J\}$, $J = \{m+1, \dots, m+p\}$; 函数 f_i ($i \in I$): $\mathbb{R}^n \rightarrow \mathbb{R}$ 称为目标函数的分量函数, c_j ($j \in J$): $\mathbb{R}^n \rightarrow \mathbb{R}$ 称为约束函数的分量函数, 并统称为分量函数.

Minimax 问题是数学规划领域中一类典型且特殊的非光滑优化问题, 它是在“最坏情形”下寻求一种“最优”的决策方案. Minimax 问题不仅广泛应用于最优控制、金融工程、统计分析、能源与环境等诸多实际问题 (参见文献 [1–5]), 而且许多数学问题可以转化为求解 Minimax 问题, 如不等式组求解问题、曲线拟和问题、 L_1 和 L_∞ 逼近问题等 (参见文献 [6, 7]). 许多学者致力于 Minimax 问题的研究, 如文献 [8–13]; 近年来的一些成果可参见文献 [14–20].

由于 Minimax 问题含极大值函数, 因此即使所有的分量函数都是光滑的 (连续可微), 该问题一般也是非光滑的 (在某些点处导数不存在, 或存在但不连续), 从而求解光滑优化问题的成熟、有效算法

英文引用格式: Jian J B, Tang C M, Tang F. A feasible descent bundle method for inequality constrained Minimax problems (in Chinese). Sci Sin Math, 2015, 45: 2001–2024, doi: 10.1360/N012015-00022

不能直接应用于求解 Minimax 问题. 然而, 若将 Minimax 问题当作一般的非光滑优化问题, 再用非光滑优化方法 (如次梯度法、束方法或梯度取样法等^[21-23]) 求解, 而忽视其特殊结构, 显然也是不理想的. 因此, 针对这类特殊非光滑优化问题, 设计出计算量小、收敛性好的数值方法, 具有重要的理论和实践意义. 下面分两种情形简要介绍求解 Minimax 问题的现有方法.

情形 I 在算法设计时, 需要利用问题 (1.1) 分量函数的一阶、甚至二阶导数, 故而需假设分量函数都是光滑的 (一阶或二阶连续可微). 针对此类情形, 主要是对原问题进行光滑化处理, 具体实施又有两种策略. 一种是通过引入光滑化函数, 用序列光滑优化问题逼近原问题, 当满足一定条件时, 可由光滑化函数得到原问题的精确解, 如文献 [8, 9]. 另一种是通过引入辅助变量, 将 Minimax 问题等价转化为一个光滑优化问题, 再针对该等价问题的特殊结构, 设计有效的数值方法, 如罚函数法^[10]、信赖域法^[11, 12]、序列二次规划法^[13, 14]、序列二次约束二次规划法^[15] 等.

情形 II 在算法设计时无需利用分量函数的导数. 在一些实际问题中, Minimax 问题的分量函数可能不可微, 或者可微但导数计算的成本较高, 甚至无法计算, 从而考虑利用函数的次梯度, 或者对梯度进行近似计算. 文献 [16] 针对分量函数是凸函数的无约束 Minimax 问题, 提出了一个增量型 (incremental) 束方法. 该方法采用分量函数的次梯度产生目标函数的割平面近似模型, 且每次迭代只需一个分量函数产生新的割平面, 从而充分利用了 Minimax 问题的特殊结构, 有利于减少计算量. 文献 [17] 将文献 [16] 的方法应用于求解整数规划的 Lagrange 对偶问题, 并获得了较好的数值效果. 文献 [18] 进一步将文献 [16] 的思想推广到求解半无限 Minimax 问题. 文献 [19, 20] 通过结合无导数技术和梯度取样技术, 提出了一类求解 Minimax 问题的无导数型梯度取样方法. 该类方法近似计算函数的梯度, 并利用梯度取样方法的思想产生近似的下降方向.

需要指出的是, 上述非光滑方法^[16, 18-20] 只能求解无约束的 Minimax 问题, 而针对含约束的情形, 目前还未查阅到相关成果. 因此, 本文的目的就是针对不等式约束 Minimax 问题 (1.1) 的分量函数可能为非光滑函数的情形, 提出一个有效的求解方法. 为了平衡目标函数的下降性和约束函数的可行性, 在非光滑领域常用的方法有罚函数法^[24, 25]、改进函数 (improvement function) 法^[21, 26]、滤子 (filter) 法^[27] 等. 由于保持有效迭代点的可行性不仅可以带来理论分析的便利, 而且有利于某些实际应用, 例如, 有些实际问题的目标函数在可行域外没有定义, 从而需要限制在可行域内计算函数值. 因此, 本文希望保持有效迭代点的可行性, 故而采取第二种方法, 即改进函数法.

具体地, 本文在假设所有的分量函数均为凸函数 (非凸的情形要复杂很多, 可作为后续工作考虑) 但不一定可微的前提下, 借助于函数的次梯度, 结合束方法思想和文献 [16] 的部分割平面模型技术, 提出一个求解问题 (1.1) 的可行下降束方法. 更直观地讲, 我们将文献 [16] 的无约束方法推广到带约束的情形. 本文算法在每次无效步迭代中, 仅需利用一个分量函数的函数值和次梯度产生新的割平面, 并且能够保证有效迭代点落在可行域内, 以及目标函数值在有效点列上具有下降性. 为了克服数值计算和存储的困难, 进一步引入次梯度聚集技术, 对束集中的次梯度进行聚集. 算法具备全局收敛性, 且初步的数值试验表明算法是有效的.

本文结构如下: 第 2 节介绍与本文密切相关的背景知识; 第 3 节给出算法及其性质; 第 4 节证明算法全局收敛性; 第 5 节利用次梯度聚集技术对算法进行改进; 第 6 和第 7 节分别给出数值试验和结语.

2 预备知识

首先对问题 (1.1) 的分量函数作如下基本假设.

假设 2.1 函数 f_i ($i \in I$) 和 c_j ($j \in J$) 均为 $\mathbb{R}^n$ 上的正常凸函数.

根据假设 2.1 易知, 问题 (1.1) 的目标函数 $f(x)$ 和约束函数 $c(x)$ 均为凸函数.

为简便, 本节暂时考虑问题 (1.1) 的无约束情形, 即

$$\min_{x \in \mathbb{R}^n} f(x). \quad (2.1)$$

函数 f 在给定点 x 处的次微分可定义为

$$\partial f(x) = \{g_f \in \mathbb{R}^n : f(y) \geq f(x) + g_f^T(y - x), \forall y \in \mathbb{R}^n\},$$

其中, $g_f \in \partial f(x)$ 称为 f 在 x 处的次梯度. 对任意给定的 $\epsilon \geq 0$, 函数 f 在 x 处的 ϵ -次微分定义为

$$\partial_\epsilon f(x) = \{g_f \in \mathbb{R}^n : f(y) \geq f(x) + g_f^T(y - x) - \epsilon, \forall y \in \mathbb{R}^n\}, \quad (2.2)$$

其中, $g_f \in \partial_\epsilon f(x)$ 称为 f 在 x 处的 ϵ -次梯度. 特别地, $\partial f(x) = \partial_0 f(x)$. 与通常的非光滑方法一样, 本文假设有一个“黑箱子”, 对问题 (1.1) 的任意一个分量函数及任意给定的点, 能计算出相应的函数值和 (任意) 一个次梯度.

以下引理刻画了极大值函数的次微分.

引理 2.1 ^[21] 设 $f(x) = \max\{f_i(x), i \in I\}$, 且 $f_i, i \in I$ 均为凸函数, 则

$$\partial f(x) = \text{conv}\{g_{f_i} \in \partial f_i(x) : i \in I(x)\},$$

其中, conv 表示集合的凸包, $I(x) = \{i \in I : f_i(x) = f(x)\}$.

2.1 经典割平面模型

设 k 为当前迭代的指标, $x^l, l = 1, \dots, k$ 为迭代点列 (x^k 为当前迭代点), $g_f^l \in \partial f(x^l)$ 为相应的 (任意一个) 次梯度, 定义函数 $f(x)$ 在 x^l 处的线性化函数为

$$\bar{f}^l(x) = f(x^l) + (g_f^l)^T(x - x^l), \quad l = 1, \dots, k. \quad (2.3)$$

割平面 (cutting planes) 法 ^[28, 29] 通过对这些线性化函数取最大来定义函数 $f(x)$ 的割平面模型:

$$\tilde{f}_{\text{cp}}^k(x) = \max_{l=1, \dots, k} \{f(x^l) + (g_f^l)^T(x - x^l)\},$$

并通过求解如下极小化问题产生新的迭代点 x^{k+1} :

$$\min_{x \in \mathbb{R}^n} \tilde{f}_{\text{cp}}^k(x). \quad (2.4)$$

由假设 2.1 易知, $\tilde{f}_{\text{cp}}^k(x) \leq f(x)$, 故其为 $f(x)$ 的下近似函数. 问题 (2.4) 可等价如下一个线性规划问题:

$$\begin{aligned} \min_{v \in \mathbb{R}, x \in \mathbb{R}^n} \quad & v \\ \text{s.t.} \quad & f(x^l) + (g_f^l)^T(x - x^l) \leq v, \quad l = 1, \dots, k. \end{aligned} \quad (2.5)$$

2.2 部分割平面模型

如果直接利用上述割平面法求解问题 (2.1), 则在每个点 x^l 处, 需要计算函数值 $f(x^l)$ 及对应的一个次梯度 $g_f^l \in \partial f(x^l)$, 从而需要计算所有分量函数 f_i ($i \in I$) 在 x^l 处的函数值及一个次梯度, 计算量相对较大. 因此, 为了避免完全计算每个分量函数 f_i 在同一点处的函数值, 受到增量型方法 (如文献 [30]) 启发, Gaudioso 等人 [16] 提出每次仅利用一个分量函数的信息产生新的割平面, 即在点 x^l 处, 从指标集 I 中选取一个指标 (至于如何选取, 将在第 3 节中讨论), 记为 i_l , 对应计算该分量函数的函数值 $f_{i_l}(x^l)$, 及相应的一个次梯度 $g_{f_{i_l}}^l \in \partial f_{i_l}(x^l)$, 进而得到该分量函数的线性化函数 (即割平面函数):

$$\bar{f}_{i_l}(x) = f_{i_l}(x^l) + (g_{f_{i_l}}^l)^T(x - x^l), \quad l = 1, \dots, k.$$

因此, 可以通过对这些线性化函数取最大来定义 $f(x)$ 的部分割平面模型:

$$\check{f}^k(x) = \max_{l=1, \dots, k} \{f_{i_l}(x^l) + (g_{f_{i_l}}^l)^T(x - x^l)\}.$$

根据假设 2.1 可知, $\check{f}^k(x) \leq f(x)$, 故其亦为 $f(x)$ 的下近似函数. 文献 [16] 基于该模型设计相应下降算法, 证明了算法的全局收敛性, 并获得了较好的数值效果.

2.3 束方法

割平面法有两个重要缺陷: 一是子问题 (2.5) 可能出现无界解; 二是随着迭代次数的增加, 子问题 (2.5) 的约束条件会不断增加, 从而导致存储和数值计算的困难. 束方法 [22] 能较好地解决这两个问题, 因此被视为稳定化的割平面法, 是目前求解一般非光滑优化问题最有效的方法之一.

为叙述简便, 以及较好地突出本文方法的思想, 此处及第 3 和第 4 节中将暂不考虑第二个缺陷, 而是将其留到第 5 节再系统描述.

束方法存储以往迭代的“一束”信息 (a bundle of information):

$$\mathcal{B}_k = \{(x^l, f(x^l), g_f^l), l = 1, \dots, k\}, \quad (y^k, f(y^k)), \quad (2.6)$$

其中, $y^k \in \{x^l\}_{l=1}^k$ 是当前 (在充分下降意义下) 使得目标函数值“最小”的点, 常称为稳定中心. $\mathcal{B}_k$ 称为束集 (或捆集 [31]), 用记号 $l \in \mathcal{B}_k$ 表示在束集 $\mathcal{B}_k$ 中存在一个标号为 l 的元素, 即 $(x^l, f(x^l), g_f^l) \in \mathcal{B}_k$ (尽管严格来说这个记号不规范, 但不会引起混淆, 而且会简化下面的表述 (参见文献 [26])).

为了克服割平面法的第一个缺陷, 利用邻近点方法思想 [32], 考虑如下极小化问题:

$$\min_{x \in \mathbb{R}^n} \check{f}_{\mathcal{B}}^k(x) + \frac{1}{2} \rho_k \|x - y^k\|^2, \quad (2.7)$$

其中, $\rho_k > 0$ 为邻近点参数, 其作用类似于信赖域法中的信赖域半径 [31], 割平面模型 $\check{f}_{\mathcal{B}}^k(x)$ 定义为

$$\check{f}_{\mathcal{B}}^k(x) = \max_{l \in \mathcal{B}_k} \{f(x^l) + (g_f^l)^T(x - x^l)\}. \quad (2.8)$$

由于此处没有考虑对 $\mathcal{B}_k$ 的元素进行选取和聚集, 因此, $\check{f}_{\mathcal{B}}^k(x) = \check{f}_{\text{cp}}^k(x)$, 但在下面小节中将会不一样.

记

$$\alpha_l^k = f(x^l) + (g_f^l)^T(y^k - x^l), \quad (2.9)$$

并令 $d = x - y^k$, 经简单计算可得

$$f(x^l) + (g_f^l)^T(x - x^l) = \alpha_l^k + (g_f^l)^T d.$$

故问题 (2.7) 可转换为求解如下搜索方向子问题:

$$\begin{aligned} \min_{v \in \mathbb{R}, d \in \mathbb{R}^n} \quad & v + \frac{1}{2} \rho_k \|d\|^2 \\ \text{s.t.} \quad & \alpha_l^k + (g_f^l)^T d \leq v, \quad l \in \mathcal{B}_k. \end{aligned} \quad (2.10)$$

显然, 子问题 (2.10) 是一个凸二次规划, 且有唯一最优解, 设其最优解为 (v_k, d^k) . 由于 $\alpha_l^k \leq f(y^k)$, 故 $(v, d) = (f(y^k), 0)$ 为 (2.10) 的可行解, 从而 $v_k \leq f(y^k)$. 若 $v_k = f(y^k)$, 则 $d^k = 0$, 算法可终止; 否则进行充分下降性测试: 令 $x^{k+1} = y^k + d^k$, $\xi \in (0, 1)$, 若

$$f(x^{k+1}) \leq f(y^k) + \xi(v_k - f(y^k)),$$

则令 $y^{k+1} = x^{k+1}$ (充分下降性成立, 更新稳定中心, 称为有效步, 并称 x^{k+1} 为有效迭代点); 否则, 令 $y^{k+1} = y^k$ (无效步, 稳定中心不变).

不管是有效步还是无效步, 均需更新束集

$$\mathcal{B}_{k+1} = \mathcal{B}_k \cup \{(x^{k+1}, f(x^{k+1})), g_f^{k+1}\},$$

其中, $g_f^{k+1} \in \partial f(x^{k+1})$. 更新束集相当于在模型 (2.8) 中加入割平面函数

$$f(x^{k+1}) + (g_f^{k+1})^T (x - x^{k+1}),$$

使模型得到改善. 特别地, 当出现无效步时, 相当于在子问题 (2.10) 中加入约束条件

$$\alpha_{k+1}^{k+1} + (g_f^{k+1})^T d \leq v,$$

并使得原来的解 (v_k, d^k) 不可行, 即被“割”掉.

3 算法及性质

本节将针对不等式约束 Minimax 问题 (1.1), 结合束方法思想和部分割平面模型技术, 提出一个可行下降束方法. 记问题 (1.1) 的可行域为

$$X = \{x \in \mathbb{R}^n : c(x) \leq 0\}.$$

引入改进函数^[21]

$$H_y(x) := \max\{f(x) - f(y), c(x)\}, \quad \forall x \in \mathbb{R}^n, \quad (3.1)$$

其中, $y \in \mathbb{R}^n$ 为相对固定的量. 当 $y \in X$ 时, 有 $H_y(y) = 0$, 若进一步满足 $H_y(x) < H_y(y) = 0$, 则有 $f(x) < f(y)$ 及 $c(x) < 0$, 这说明, x 不仅是可行点, 而且其函数值比 y 的小. 所以, x 是比 y 改进的点, 由此称函数 $H_y(x)$ 为问题 (1.1) 的改进函数.

由引理 2.1 可知, 函数 $H_y(\cdot)$ 在 x 处的次微分可表述为

$$\partial H_y(x) = \begin{cases} \partial f(x), & \text{若 } f(x) - f(y) > c(x), \\ \text{conv}\{\partial f(x) \cup \partial c(x)\}, & \text{若 } f(x) - f(y) = c(x), \\ \partial c(x), & \text{若 } f(x) - f(y) < c(x). \end{cases}$$

以下引理刻画了改进函数 (3.1) 的重要性质, 其证明可参见文献 [21].

引理 3.1 设假设 2.1 成立, 且问题 (1.1) 满足 Slater 约束规格, 即存在一个点 $\tilde{x} \in \mathbb{R}^n$ 使得 $c(\tilde{x}) < 0$, 则下面三个结论等价:

- (1) $\bar{y}$ 是问题 (1.1) 的最优解;
- (2) $\min\{H_{\bar{y}}(x) : x \in \mathbb{R}^n\} = H_{\bar{y}}(\bar{y}) = 0$;
- (3) $0 \in \partial H_{\bar{y}}(\bar{y})$.

3.1 约束 Minimax 问题的部分割平面模型

本小节将利用第 2.2 小节的部分割平面模型技术, 提出适合约束 Minimax 问题的部分割平面模型, 进而导出本文所采用的搜索方向子问题.

设 $y^k \in X$ 为当前的稳定中心, 引入记号

$$\begin{aligned}\beta_l^k &:= \bar{f}_{i_l}(y^k) = f_{i_l}(x^l) + (g_{f_{i_l}}^l)^T(y^k - x^l), \quad l = 1, \dots, k, \\ \gamma_l^k &:= \bar{c}_{j_l}(y^k) = c_{j_l}(x^l) + (g_{c_{j_l}}^l)^T(y^k - x^l), \quad l = 1, \dots, k,\end{aligned}$$

其中, i_l 和 j_l 分别为指标集 I 和 J 中的某个指标, 且依赖于点 x^l ; $g_{f_{i_l}}^l \in \partial f_{i_l}(x^l)$, $g_{c_{j_l}}^l \in \partial c_{j_l}(x^l)$. 直观地看, β_l^k 和 γ_l^k 分别为目标和约束函数的某个分量的线性化函数在 y^k 处的值. 根据假设 2.1 可知,

$$\beta_l^k \leq f_{i_l}(y^k) \leq f(y^k), \quad \gamma_l^k \leq c_{j_l}(y^k) \leq c(y^k) \leq 0. \quad (3.2)$$

选取束集 $\mathcal{B}_k^f$ 和 $\mathcal{B}_k^c$ 满足

$$\begin{aligned}\mathcal{B}_k^f &\subseteq \{(x^l, f_{i_l}(x^l), g_{f_{i_l}}^l, \beta_l^k), l = 1, \dots, k\}, \\ \mathcal{B}_k^c &\subseteq \{(x^l, c_{j_l}(x^l), g_{c_{j_l}}^l, \gamma_l^k), l = 1, \dots, k\},\end{aligned}$$

并分别作为问题 (1.1) 的目标函数和约束函数的束集, 具体如何产生将在下面算法中给出, 再记

$$\mathcal{B}_k = \mathcal{B}_k^f \cup \mathcal{B}_k^c.$$

根据第 2.2 小节中的部分割平面模型思想, 并利用束集 $\mathcal{B}_k^f$ 和 $\mathcal{B}_k^c$ 中的信息, 分别定义函数 $f(x)$ 和 $c(x)$ 的部分割平面模型如下:

$$\check{f}_{\mathcal{B}}^k(x) = \max_{l \in \mathcal{B}_k^f} \{f_{i_l}(x^l) + (g_{f_{i_l}}^l)^T(x - x^l)\} = \max_{l \in \mathcal{B}_k^f} \{\beta_l^k + (g_{f_{i_l}}^l)^T(x - y^k)\}, \quad (3.3)$$

$$\check{c}_{\mathcal{B}}^k(x) = \max_{l \in \mathcal{B}_k^c} \{c_{j_l}(x^l) + (g_{c_{j_l}}^l)^T(x - x^l)\} = \max_{l \in \mathcal{B}_k^c} \{\gamma_l^k + (g_{c_{j_l}}^l)^T(x - y^k)\}. \quad (3.4)$$

由此, 可进一步定义改进函数 $H_{y^k}(x)$ 的部分割平面函数为

$$\check{H}_{y^k}(x) = \max\{\check{f}_{\mathcal{B}}^k(x) - f(y^k), \check{c}_{\mathcal{B}}^k(x)\}. \quad (3.5)$$

由假设 2.1 可知,

$$\check{f}_{\mathcal{B}}^k(x) \leq f(x), \quad \check{c}_{\mathcal{B}}^k(x) \leq c(x), \quad \check{H}_{y^k}(x) \leq H_{y^k}(x).$$

因此, $\check{f}_{\mathcal{B}}^k(x)$, $\check{c}_{\mathcal{B}}^k(x)$ 和 $\check{H}_{y^k}(x)$ 分别为 $f(x)$, $c(x)$ 和 $H_{y^k}(x)$ 的下近似函数.

根据引理 3.1 及邻近点束方法思想^[32], 本文通过求解如下问题获得新迭代点 x^{k+1} :

$$\min_{x \in \mathbb{R}^n} \check{H}_{y^k}(x) + \frac{1}{2} \rho_k \|x - y^k\|^2. \quad (3.6)$$

根据 (3.3) 和 (3.4), 并令 $d = x - y^k$, 问题 (3.6) 可转化为求解如下二次规划 (QP) 子问题:

$$\begin{aligned} \min_{z \in \mathbb{R}, d \in \mathbb{R}^n} \quad & z + \frac{1}{2} \rho_k \|d\|^2 \\ \text{QP}(\mathcal{B}_k) \quad & \text{s.t.} \quad \beta_l^k - f(y^k) + (g_{f_{i_l}}^l)^T d \leq z, \quad l \in \mathcal{B}_k^f, \\ & \gamma_l^k + (g_{c_{j_l}}^l)^T d \leq z, \quad l \in \mathcal{B}_k^c. \end{aligned} \quad (3.7)$$

由 (3.2) 可知, $(0, 0)$ 是 $\text{QP}(\mathcal{B}_k)$ 的一个可行解, 故由文献 [33, 推论 3.4.5] 可知, 子问题 $\text{QP}(\mathcal{B}_k)$ 有唯一最优解, 设其为 (z_k, d^k) . 进而结合 $(0, 0)$ 是可行解, 有

$$z_k + \frac{1}{2} \rho_k \|d^k\|^2 \leq 0, \quad z_k \leq -\frac{1}{2} \rho_k \|d^k\|^2 \leq 0. \quad (3.8)$$

子问题 $\text{QP}(\mathcal{B}_k)$ 的 Lagrange 对偶问题为

$$\begin{aligned} \min_{\lambda \in \mathbb{R}^{|\mathcal{B}_k^f|}, \mu \in \mathbb{R}^{|\mathcal{B}_k^c|}} \quad & \frac{1}{2\rho_k} \left\| \sum_{l \in \mathcal{B}_k^f} \lambda_l g_{f_{i_l}}^l + \sum_{l \in \mathcal{B}_k^c} \mu_l g_{c_{j_l}}^l \right\|^2 - \sum_{l \in \mathcal{B}_k^f} \lambda_l (\beta_l^k - f(y^k)) - \sum_{l \in \mathcal{B}_k^c} \mu_l \gamma_l^k \\ \text{DP}(\mathcal{B}_k) \quad & \text{s.t.} \quad \lambda_l \geq 0, \quad l \in \mathcal{B}_k^f, \quad \mu_l \geq 0, \quad l \in \mathcal{B}_k^c, \\ & \sum_{l \in \mathcal{B}_k^f} \lambda_l + \sum_{l \in \mathcal{B}_k^c} \mu_l = 1. \end{aligned}$$

设 (λ^k, μ^k) 为子问题 $\text{DP}(\mathcal{B}_k)$ 的最优解, 易知其亦为 $\text{QP}(\mathcal{B}_k)$ 的 KKT 乘子, 且如下关系成立:

$$d^k = -\frac{1}{\rho_k} g^k, \quad z_k = -\frac{1}{\rho_k} \|g^k\|^2 - \epsilon_k, \quad (3.9)$$

其中,

$$g^k = \sum_{l \in \mathcal{B}_k^f} \lambda_l^k g_{f_{i_l}}^l + \sum_{l \in \mathcal{B}_k^c} \mu_l^k g_{c_{j_l}}^l, \quad \epsilon_k = \sum_{l \in \mathcal{B}_k^f} \lambda_l^k (f(y^k) - \beta_l^k) - \sum_{l \in \mathcal{B}_k^c} \mu_l^k \gamma_l^k. \quad (3.10)$$

以下引理将用于刻画问题 (1.1) 的近似最优性条件.

引理 3.2 向量 g^k 是改进函数 $H_{y^k}(\cdot)$ 在 y^k 处的 ϵ_k -次梯度, 即 $g^k \in \partial_{\epsilon_k} H_{y^k}(y^k)$.

证明 根据 $f_i, i \in I$ 的凸性, 对任意的 $x \in \mathbb{R}^n$, 有

$$\begin{aligned} f(x) &\geq f_{i_l}(x) \geq f_{i_l}(x^l) + (g_{f_{i_l}}^l)^T (x - x^l) \\ &= f(y^k) + (g_{f_{i_l}}^l)^T (x - y^k) - (f(y^k) - \beta_l^k), \quad l \in \mathcal{B}_k^f. \end{aligned} \quad (3.11)$$

在 (3.11) 两端同乘以 $\lambda_l^k (\geq 0)$, 再对 $l \in \mathcal{B}_k^f$ 相加, 得

$$\sum_{l \in \mathcal{B}_k^f} \lambda_l^k (f(x) - f(y^k)) \geq \sum_{l \in \mathcal{B}_k^f} \lambda_l^k (g_{f_{i_l}}^l)^T (x - y^k) - \sum_{l \in \mathcal{B}_k^f} \lambda_l^k (f(y^k) - \beta_l^k). \quad (3.12)$$

类似地, 根据 $c_j, j \in J$ 的凸性, 有

$$\begin{aligned} c(x) &\geq c_{j_l}(x) \geq c_{j_l}(x^l) + (g_{c_{j_l}}^l)^T (x - x^l) \\ &= (g_{c_{j_l}}^l)^T (x - y^k) + \gamma_l^k. \end{aligned}$$

因此, 结合 $\mu_l^k \geq 0, l \in \mathcal{B}_k^c$ 可得

$$\sum_{l \in \mathcal{B}_k^c} \mu_l^k c(x) \geq \sum_{l \in \mathcal{B}_k^c} \mu_l^k (g_{c_{j_l}}^l)^T (x - y^k) + \sum_{l \in \mathcal{B}_k^c} \mu_l^k \gamma_l^k. \quad (3.13)$$

由于 $\sum_{l \in \mathcal{B}_k^f} \lambda_l^k + \sum_{l \in \mathcal{B}_k^c} \mu_l^k = 1$, 结合 (3.12)、(3.13) 和 (3.10), 以及 $H_{y^k}(y^k) = 0$, 有

$$\begin{aligned} H_{y^k}(x) &= \max\{f(x) - f(y^k), c(x)\} \geq \sum_{l \in \mathcal{B}_k^f} \lambda_l^k (f(x) - f(y^k)) + \sum_{l \in \mathcal{B}_k^c} \mu_l^k c(x) \\ &\geq \sum_{l \in \mathcal{B}_k^f} \lambda_l^k (g_{f_{i_l}}^l)^T (x - y^k) - \sum_{l \in \mathcal{B}_k^f} \lambda_l^k (f(y^k) - \beta_l^k) + \sum_{l \in \mathcal{B}_k^c} \mu_l^k (g_{c_{j_l}}^l)^T (x - y^k) + \sum_{l \in \mathcal{B}_k^c} \mu_l^k \gamma_l^k \\ &= H_{y^k}(y^k) + (g^k)^T (x - y^k) - \epsilon_k, \quad \forall x \in \mathbb{R}^n. \end{aligned}$$

上式结合定义 (2.2) 可得引理成立. $\square$

由引理 3.1 和 3.2 可知, 当 $\|g^k\|$ 和 ϵ_k 充分小时, y^k 可作为问题 (1.1) 的一个近似最优解.

3.2 束集更新策略

在算法迭代开始前, 需要初始化束集. 选取初始点 $x^1 \in X$, 置初始稳定中心 $y^1 = x^1$, 选取指标 $i_1 \in I$, $j_1 \in J$, 初始化束集:

$$\mathcal{B}_1^f = \{(x^1, f_{i_1}(x^1), g_{f_{i_1}}^1, \beta_1^1)\}, \quad \mathcal{B}_1^c = \{(x^1, c_{j_1}(x^1), g_{c_{j_1}}^1, \gamma_1^1)\}, \quad (3.14)$$

其中,

$$g_{f_{i_1}}^1 \in \partial f_{i_1}(x^1), \quad \beta_1^1 = f_{i_1}(x^1), \quad g_{c_{j_1}}^1 \in \partial c_{j_1}(x^1), \quad \gamma_1^1 = c_{j_1}(x^1).$$

此处, 由于在 $\text{QP}(\mathcal{B}_1)$ 中需要计算 $f(y^1)$, 即 $f(x^1)$, 因此, 根据引理 2.1, 可以选取 $i_1 \in I(x^1)$, 于是便有

$$g_{f_{i_1}}^1 \in \partial f(x^1), \quad f_{i_1}(x^1) = f(x^1).$$

从数值计算角度来看, 这样可能会更好.

算法在迭代的过程中, 需要不断地对束集进行更新, 进而利用束集的信息产生新的搜索方向子问题. 下面分三种情形讨论束集的更新策略.

(1) 束集重置.

直观地看, $\|g^k\|$ 可视为一种“最优性”的衡量, 而 ϵ_k 可视为“模型误差”的衡量. 因此, 当 $\|g^k\|$ “较小”, 而 ϵ_k “较大”时 (详见算法第 1 和第 2 步), 可认为当前的模型较差, 故而不宜再用现有的迭代点列构造割平面模型. 此处考虑仅保留当前的稳定中心, 而删除其他所有的点, 即令 $y^{k+1} = y^k$, 并重置束集:

$$\mathcal{B}_{k+1}^f = \{(x^{l(k+1)}, f(x^{l(k+1)}), g_f^{l(k+1)}, \beta_{l(k+1)}^{k+1})\}, \quad (3.15)$$

$$\mathcal{B}_{k+1}^c = \begin{cases} \{(x^{l(k+1)}, c(x^{l(k+1)}), g_c^{l(k+1)}, \gamma_{l(k+1)}^{k+1})\}, & \text{若 } c(x^{l(k+1)}) = 0, \\ \emptyset, & \text{若 } c(x^{l(k+1)}) < 0, \end{cases} \quad (3.16)$$

其中, $l(k+1)$ 是 $\{1, \dots, k+1\}$ 中使得 $y^{k+1} = x^{l(k+1)}$ 的指标, 即 $x^{l(k+1)}$ 是与当前稳定中心对应的迭代点;

$$\begin{aligned} g_f^{l(k+1)} &\in \partial f(x^{l(k+1)}), \quad \beta_{l(k+1)}^{k+1} = f(x^{l(k+1)}), \\ g_c^{l(k+1)} &\in \partial c(x^{l(k+1)}), \quad \gamma_{l(k+1)}^{k+1} = c(x^{l(k+1)}). \end{aligned}$$

在 (3.16) 中, 当 $c(x^{l(k+1)}) < 0$ 时, 说明 $x^{l(k+1)}$ 是一个严格内点, 故仅考虑对目标函数增加割平面, 而约束函数不加, 即令 $\mathcal{B}_{k+1}^c = \emptyset$, 这同时也是重置策略能够有限次终止的一个保证 (见引理 4.1).

(2) 束集扩充.

如果找到一个指标 $h \in I$, 使得分量函数 f_h 在试探点 $x^{k+1} := y^k + d^k$ 处 “下降不充分”, 即

$$f_h(x^{k+1}) - f(y^k) > z_k + \delta_k,$$

则视当前迭代为无效步, 令 $y^{k+1} = y^k$, 并利用该函数的信息产生新的割平面, 即扩充束集:

$$\mathcal{B}_{k+1}^f = \mathcal{B}_k^f \cup \{(x^{k+1}, f_h(x^{k+1}), g_{f_h}^{k+1}, \beta_{k+1}^{k+1})\}, \quad \mathcal{B}_{k+1}^c = \mathcal{B}_k^c, \quad (3.17)$$

其中, 参数 $\delta_k > 0$,

$$g_{f_h}^{k+1} \in \partial f_h(x^{k+1}), \quad \beta_{k+1}^{k+1} = f_h(x^{k+1}) + (g_{f_h}^{k+1})^T (y^{k+1} - x^{k+1}).$$

类似地, 若找到一个指标 $h \in J$, 使得分量函数 c_h 在 x^{k+1} 处不可行, 即 $c_h(x^{k+1}) > 0$, 则视当前迭代为无效步, 令 $y^{k+1} = y^k$, 并扩充束集:

$$\mathcal{B}_{k+1}^c = \mathcal{B}_k^c \cup \{(x^{k+1}, c_h(x^{k+1}), g_{c_h}^{k+1}, \gamma_{k+1}^{k+1})\}, \quad \mathcal{B}_{k+1}^f = \mathcal{B}_k^f, \quad (3.18)$$

其中,

$$g_{c_h}^{k+1} \in \partial c_h(x^{k+1}), \quad \gamma_{k+1}^{k+1} = c_h(x^{k+1}) + (g_{c_h}^{k+1})^T (y^{k+1} - x^{k+1}).$$

值得指出的是, 在新的试探点 x^{k+1} 处, 仅新增一个割平面, 要么针对目标函数加, 要么针对约束函数加. 如果两个同时加, 会导致理论分析的困难 (见引理 4.1). 在束集扩充的过程中会不断强化新迭代点的可行性和下降性, 直到完全满足 (见引理 3.4).

(3) 束集更新并扩充.

当算法产生有效步, 稳定中心更新时 (即 $y^{k+1} = x^{k+1}$), 需对束集中的元素 β_l^k 和 γ_l^k 进行更新, 即令

$$\beta_l^{k+1} = f_{i_l}(x^l) + (g_{f_{i_l}}^l)^T (y^{k+1} - x^l) = \beta_l^k + (g_{f_{i_l}}^l)^T (y^{k+1} - y^k), \quad l \in \mathcal{B}_k^f, \quad (3.19)$$

$$\gamma_l^{k+1} = c_{j_l}(x^l) + (g_{c_{j_l}}^l)^T (y^{k+1} - x^l) = \gamma_l^k + (g_{c_{j_l}}^l)^T (y^{k+1} - y^k), \quad l \in \mathcal{B}_k^c. \quad (3.20)$$

然后, 在新的迭代点 x^{k+1} 处产生新的割平面, 即扩充束集:

$$\begin{aligned} \mathcal{B}_{k+1}^f &= \{(x^l, f_{i_l}(x^l), g_{f_{i_l}}^l, \beta_l^{k+1}), l \in \mathcal{B}_k^f\} \cup \{(x^{k+1}, f(x^{k+1}), g_f^{k+1}, \beta_{k+1}^{k+1})\}, \\ \mathcal{B}_{k+1}^c &= \{(x^l, c_{j_l}(x^l), g_{c_{j_l}}^l, \gamma_l^{k+1}), l \in \mathcal{B}_k^c\} \cup \{(x^{k+1}, c(x^{k+1}), g_c^{k+1}, \gamma_{k+1}^{k+1})\}, \end{aligned} \quad (3.21)$$

其中

$$g_f^{k+1} \in \partial f(x^{k+1}), \quad \beta_{k+1}^{k+1} = f(x^{k+1}), \quad g_c^{k+1} \in \partial c(x^{k+1}), \quad \gamma_{k+1}^{k+1} = c(x^{k+1}).$$

注 3.1 (1) 由于当稳定中心更新时, 已经隐含着计算出了所有分量函数在稳定中心 y^{k+1} 处的值, 进而知道 $f(y^{k+1})$ 和 $c(y^{k+1})$, 因此, (3.15) 和 (3.16) 的束集重置以及 (3.21) 的扩充, 均不会额外增加函数值的计算量, 而且根据引理 2.1, 只需要分别计算一个目标和约束函数的分量函数的次梯度. 此外, 在子问题 $\text{QP}(\mathcal{B}_{k+1})$ 中用到的 $f(y^{k+1})$ 也不用重新计算.

(2) 根据 (3.19) 和 (3.20) 中元素 β_l^k 和 γ_l^k 的更新方式, 以及子问题 $\text{QP}(\mathcal{B}_k)$ 的结构, 在实际计算中并不需要存储过往的迭代点 $\{x^l\}$ 及相应的函数值, 但为了形式上的完整性, 我们仍在束集中保留它们.

以下引理指出, 当根据 (3.17) 或 (3.18) 扩充束集时, 会产生一个 (显著) 更好的部分割平面模型, 即 $\text{QP}(\mathcal{B}_k)$ 的解 (z_k, d^k) 不再是新子问题 $\text{QP}(\mathcal{B}_{k+1})$ 的可行解. 更直观地讲, (z_k, d^k) 被新加入的割平面 “割掉”, 故而称之为割性质.

引理 3.3 (割性质) 设 $x^{k+1} = y^k + d^k$, $y^{k+1} = y^k$ (无效步, 稳定中心不变), 给定参数 $\delta_k > 0$, 假设下面两种情形之一成立:

- (1) 若存在指标 $h \in I$, 使得 $f_h(x^{k+1}) - f(y^k) > z_k + \delta_k$, 且利用 (3.17) 扩充束集;
- (2) 若存在指标 $h \in J$, 使得 $c_h(x^{k+1}) > 0$, 且利用 (3.18) 扩充束集,

则 (z_k, d^k) 不是 $\text{QP}(\mathcal{B}_{k+1})$ 的可行解.

证明 若情形 (1) 成立, 则

$$\begin{aligned} (g_{f_h}^{k+1})^T d^k + \beta_{k+1}^{k+1} - f(y^k) &= (g_{f_h}^{k+1})^T d^k + f_h(x^{k+1}) - (g_{f_h}^{k+1})^T d^k - f(y^k) \\ &= f_h(x^{k+1}) - f(y^k) > z_k + \delta_k > z_k. \end{aligned}$$

若情形 (2) 成立, 则

$$\begin{aligned} \gamma_{k+1}^{k+1} + (g_{c_h}^{k+1})^T d^k &= c_h(x^{k+1}) - (g_{c_h}^{k+1})^T d^k + (g_{c_h}^{k+1})^T d^k \\ &> 0 \geq z_k. \end{aligned}$$

因此, 不管出现哪种情形, (z_k, d^k) 均不是 $\text{QP}(\mathcal{B}_{k+1})$ 的可行解. □

3.3 约束 Minimax 问题的可行下降束方法

基于以上分析, 下面给出本文算法的具体步骤.

算法 3.1 (约束 Minimax 问题的可行下降束方法)

第 0 步 选取初始 (有效) 点 $x^1 \in X$, 令 $y^1 = x^1$, 根据 (3.14) 初始化束集. 选取参数 $\eta > 0$, $\rho_1 = \bar{\rho} > 0$ 和 $\delta_1 = \bar{\delta} > 0$ 满足 $\eta^2 > \bar{\rho}\bar{\delta}$, 以及 $\epsilon > 0$, $\sigma > 1$. 令 $W = I \cup J$, $k = 1$.

第 1 步 求解子问题 $\text{QP}(\mathcal{B}_k)$ 或 $\text{DP}(\mathcal{B}_k)$ 得最优解 (z_k, d^k) 和 (λ^k, μ^k) . 令 $x^{k+1} = y^k + d^k$. 若 $\|g^k\| > \eta$, 进入第 3 步.

第 2 步 若 $\epsilon_k \leq \epsilon$, 则算法终止 (满足近似最优性条件); 否则, 令 $y^{k+1} = y^k$, 并根据 (3.15) 和 (3.16) 重置束集, 令 $\rho_{k+1} = \sigma\rho_k$, $\delta_{k+1} = \delta_k/\sigma$, $k := k + 1$, 返回第 1 步.

第 3 步 选取一个指标 $h \in W$.

(1) 若 $h \in I$, 且 $f_h(x^{k+1}) - f(y^k) > z_k + \delta_k$, 则令 $y^{k+1} = y^k$ (无效步), 并根据 (3.17) 扩充束集. 恢复指标集 $W = I \cup J$, 令 $\rho_{k+1} = \rho_k$, $\delta_{k+1} = \delta_k$, $k := k + 1$, 返回第 1 步.

(2) 若 $h \in J$, 且 $c_h(x^{k+1}) > 0$, 则令 $y^{k+1} = y^k$ (无效步), 并根据 (3.18) 扩充束集. 恢复指标集 $W = I \cup J$, 令 $\rho_{k+1} = \rho_k$, $\delta_{k+1} = \delta_k$, $k := k + 1$, 返回第 1 步.

第 4 步 令 $W := W \setminus \{h\}$. 若 $W \neq \emptyset$, 返回第 3 步.

第 5 步 令 $y^{k+1} = x^{k+1}$ (有效步, 更新稳定中心, x^{k+1} 为新的有效点), 根据 (3.21) 更新并扩充束集. 恢复指标集 $W = I \cup J$, 令 $\rho_{k+1} = \bar{\rho}$, $\delta_{k+1} = \bar{\delta}$, $k := k + 1$, 返回第 1 步.

注 3.2 (1) 如果算法 3.1 在第 2 步终止, 则满足近似最优性条件

$$\|g^k\| \leq \eta, \quad \epsilon_k \leq \epsilon.$$

从而, 由引理 3.2 可知, 当前稳定中心 y^k 是问题 (1.1) 的近似最优解.

(2) 理论上, 子问题 $QP(\mathcal{B}_k)$ 与其对偶问题 $DP(\mathcal{B}_k)$ 的解可以相互转化, 即只要求解其中一个问题, 就能简单获得另外一个问题的解. 然而, 在数值计算上, 两者之间存在一定的差异. 最典型的表现是, 当问题 (1.1) 的变量维数 n 较大时, 相应的 $QP(\mathcal{B}_k)$ 子问题的变量维数也较大, 从而求解相对困难. 本文第 5 节通过引入次梯度聚集技术, 能控制 $QP(\mathcal{B}_k)$ 的约束个数在一个 (较少的) 合理范围, 从而 $DP(\mathcal{B}_k)$ 的变量维数较小, 加之其约束较简单, 且有专门的求解方法^[34], 因此, $DP(\mathcal{B}_k)$ 比 $QP(\mathcal{B}_k)$ 更易求解.

(3) 在 $QP(\mathcal{B}_k)$ 或 $DP(\mathcal{B}_k)$ 子问题中, 表面上看需要先计算 $f(y^k)$ 的值, 而事实上, 每当算法进入第 5 步更新稳定中心之时, 已经隐含计算出了所有目标和约束函数的分量函数在新稳定中心处的值, 故此处没有额外增加计算量.

(4) 若算法在第 2 步中需要对束集进行重置, 则如第 3.2 小节中所述, 当前的部分割平面模型较差, 因此, 我们需要变得更“保守”, 即希望新的试探点更加靠近当前稳定中心, 所以需要增大邻近点参数 ρ , 即令 $\rho_{k+1} = \sigma\rho_k$. 与此同时, 为了保持 $\rho_k\delta_k < \eta^2$ (见引理 3.4), 需要减小参数 δ , 即令

$$\delta_{k+1} = \frac{\delta_k}{\sigma}.$$

(5) 在第 3 步中, 指标 $h \in W$ 的选取方式对理论分析没有任何影响. 在数值计算中, 可根据实际情形采用适当的选取方式, 如顺序选取、随机选取或根据子问题 $QP(\mathcal{B}_k)$ 的积极集信息选取等 (参见文献 [16, 30]). 此处, 作为本文方法的特点之一, 如果目标函数在可行域外没有定义, 或者有定义但计算量较大, 可以先选取约束函数指标集 J 中的指标测试可行性, 待试探点可行之后, 再选取 I 中的指标测试下降性. 显然, 不同的选取方式会产生不同的数值效果.

(6) 当算法进入第 5 步后, 产生的新迭代点 x^{k+1} (亦为新的稳定中心 y^{k+1}) 满足可行下降性 (详见下述引理 3.4), 这说明, 新的稳定中心 y^{k+1} 显著优于当前的稳定中心 y^k .

引理 3.4 (可行下降性) 若算法 3.1 进入第 5 步, 则 $c(y^{k+1}) \leq 0$, 且

$$f(y^{k+1}) - f(y^k) \leq -\frac{1}{\rho_k} \|g^k\|^2 + \delta_k < 0.$$

证明 当算法 3.1 进入第 5 步时, 必定成立下式:

$$\|g^k\| > \eta, \quad (3.22)$$

$$f_i(y^{k+1}) - f(y^k) \leq z_k + \delta_k, \quad i \in I, \quad (3.23)$$

$$c_j(y^{k+1}) \leq 0, \quad j \in J. \quad (3.24)$$

由 (3.24) 易知,

$$c(y^{k+1}) \leq 0.$$

又由 (3.23) 和 (3.9) 可得

$$\begin{aligned} f(y^{k+1}) - f(y^k) &= \max_{i \in I} \{f_i(y^{k+1}) - f(y^k)\} \leq z_k + \delta_k \\ &= -\frac{1}{\rho_k} \|g^k\|^2 - \epsilon_k + \delta_k \leq -\frac{1}{\rho_k} \|g^k\|^2 + \delta_k. \end{aligned} \quad (3.25)$$

此外, 由第 2 步可知,

$$\rho_k\delta_k = \bar{\rho}\bar{\delta} < \eta^2,$$

此结合 (3.22) 和 (3.25) 得

$$f(y^{k+1}) - f(y^k) \leq -\frac{1}{\rho_k} \|g^k\|^2 + \delta_k = -\frac{1}{\rho_k} (\|g^k\|^2 - \rho_k \delta_k) < 0.$$

证毕. $\square$

4 全局收敛性

本节将借鉴文献 [16] 的 (部分) 分析思路, 证明算法 3.1 的全局收敛性. 与文献 [16] 相比, 本文的主要不同之处在于, 由于引入了约束条件, 在引理 4.1 和 4.3 的相关证明中要兼顾目标和约束两种情形. 此外, 引理 4.1 给出了束集重置次数的一个量化的上界, 而文献 [16] 仅笼统地说存在一个上界, 但没有明确给出.

首先作如下基本假设.

假设 4.1 目标函数 $f(x)$ 在可行域 X 上有下界.

假设 4.2 算法 3.1 产生的迭代点列 $\{x^k\}$ 有界.

事实上, 假设 4.2 可替换为一个与算法迭代无关 (仅与初始点有关) 的假设: 假设目标函数的水平集交上可行域 X 有界, 即集合 $\{x : f(x) \leq f(x^1)\} \cap X$ 有界, 仅需对相应分析作适当修改. 该假设主要保证算法的重置次数有一个与 k 无关的上界 (见引理 4.1).

由假设 2.1 和 4.2, 并结合文献 [21, 引理 1.2.2] 可知, 存在常数 $L > 0$ 使得

$$\|g_f^k\| \leq L, \quad \|g_{f_i}^k\| \leq L, \quad \forall g_f^k \in \partial f(x^k), \quad \forall g_{f_i}^k \in \partial f_i(x^k), \quad \forall k, \quad \forall i \in I. \quad (4.1)$$

由算法 3.1 的迭代步骤可知, 算法在迭代过程中必定出现以下两种情形之一:

- (1) 算法在第 1 至第 4 步之间循环, 稳定中心不变, 并依次产生无效步;
- (2) 算法进入第 5 步, 完成一次有效迭代, 并更新稳定中心.

下面将证明算法 3.1 的适定性, 即算法在第 1 至第 4 步之间经过有限次循环之后, 必会进入第 5 步. 具体地, 我们将在稳定中心不变的情形下 (若发生改变, 说明已进入第 5 步), 证明以下三个结论:

- (a) 算法有限次通过第 2 步;
- (b) 算法有限次进入第 4 步;
- (c) 算法在第 1 与第 3 步之间的循环是有限的.

首先证明结论 (a).

引理 4.1 当稳定中心不变时 (即算法不进入第 5 步), 算法有限次通过第 2 步. 具体地, 若记 $x^{\bar{k}}$ 为当前有效点, $y^{\bar{k}} = x^{\bar{k}}$ 为当前稳定中心, $t_{\bar{k}}$ 为算法进入第 5 步之前通过第 2 步的次数, 则

$$t_{\bar{k}} \leq \left\lceil \frac{\ln \frac{4L^2}{\rho\epsilon}}{\ln \sigma} \right\rceil,$$

其中, 符号 $\lceil \cdot \rceil$ 表示向上取整, 即 $\lceil a \rceil$ 表示不小于 a 的最小整数.

证明 用 $k_r (\geq \bar{k})$ 表示第 r 次进入第 2 步时对应的迭代指标. 进入第 2 步之后, 将会重置束集为

$$\mathcal{B}_{k_r+1}^f = \{(x^{\bar{k}}, f(x^{\bar{k}}), g_f^{\bar{k}}, \beta_{\bar{k}}^{k_r+1})\}, \quad (4.2)$$

$$\mathcal{B}_{k_r+1}^c = \begin{cases} \{(x^{\bar{k}}, c(x^{\bar{k}}), g_c^{\bar{k}}, \gamma_{\bar{k}}^{k_r+1})\}, & \text{若 } c(x^{\bar{k}}) = 0, \\ \emptyset, & \text{若 } c(x^{\bar{k}}) < 0, \end{cases} \quad (4.3)$$

其中,

$$g_f^{\bar{k}} \in \partial f(x^{\bar{k}}), \quad \beta_k^{k_r+1} = f(x^{\bar{k}}), \quad g_c^{\bar{k}} \in \partial c(x^{\bar{k}}), \quad \gamma_k^{k_r+1} = c(x^{\bar{k}}).$$

并更新参数

$$\rho_{k_r+1} = \bar{\rho}\sigma^r, \quad \delta_{k_r+1} = \frac{\bar{\delta}}{\sigma^r}. \quad (4.4)$$

当 $k_r + 1 \leq k < k_{r+1}$ 时, 束集的更新必为以下两种情形之一:

情形 A 对某个 $h \in I$, $f_h(x^{k+1}) - f(y^k) > z_k + \delta_k$, 束集由 (3.17) 更新;

情形 B 对某个 $h \in J$, $c_h(x^{k+1}) > 0$, 束集由 (3.18) 更新.

设 (z_k, d^k) 和 (λ^k, μ^k) 分别为 $\text{QP}(\mathcal{B}_k)$ 和 $\text{DP}(\mathcal{B}_k)$ 的最优解. 当 $k_r + 1 \leq k < k_{r+1}$ 时, 根据 (4.2) 和 (4.4) 可得

$$-\frac{1}{2}\bar{\rho}\sigma^r\|d^k\|^2 \geq z_k \geq \beta_k^{k_r+1} - f(y^{\bar{k}}) + (g_f^{\bar{k}})^T d^k \geq -\|g_f^{\bar{k}}\|\|d^k\|.$$

此式结合 (4.1) 有

$$\|d^k\| \leq \frac{2}{\bar{\rho}\sigma^r}\|g_f^{\bar{k}}\| \leq \frac{2L}{\bar{\rho}\sigma^r}, \quad (4.5)$$

并进一步可得

$$z_k \geq -\frac{2L^2}{\bar{\rho}\sigma^r}. \quad (4.6)$$

对于情形 A, 结合 (4.1)、(4.5) 和 (4.6) 可知,

$$\begin{aligned} \beta_{k+1}^{k+1} - f(y^{k+1}) &= f_h(x^{k+1}) + (g_{f_h}^{k+1})^T (y^{k+1} - x^{k+1}) - f(y^{k+1}) \\ &> z_k + \delta_k - \|g_{f_h}^{k+1}\|\|d^k\| \\ &\geq \delta_k - \frac{4L^2}{\bar{\rho}\sigma^r} > -\frac{4L^2}{\bar{\rho}\sigma^r}, \quad k_r + 1 \leq k < k_{r+1}. \end{aligned} \quad (4.7)$$

设 (z_{k+1}, d^{k+1}) 和 $(\lambda^{k+1}, \mu^{k+1})$ 分别为 $\text{QP}(\mathcal{B}_{k+1})$ 和 $\text{DP}(\mathcal{B}_{k+1})$ 的最优解, 由 (4.2) 和 (4.7) 有

$$\sum_{l \in \mathcal{B}_{k+1}^f} \lambda_l^{k+1} (f(y^{k+1}) - \beta_l^{k+1}) < \frac{4L^2}{\bar{\rho}\sigma^r} \sum_{l \in \mathcal{B}_{k+1}^f} \lambda_l^{k+1}, \quad k_r \leq k < k_{r+1}. \quad (4.8)$$

类似地, 对于情形 B, 当 $k_r + 1 \leq k < k_{r+1}$ 时, 有

$$\begin{aligned} \gamma_{k+1}^{k+1} &= c_h(x^{k+1}) + (g_{c_h}^{k+1})^T (y^{k+1} - x^{k+1}) \\ &> -\frac{2L^2}{\bar{\rho}\sigma^r} > -\frac{4L^2}{\bar{\rho}\sigma^r}. \end{aligned}$$

此式结合 (4.3) 说明

$$-\sum_{l \in \mathcal{B}_{k+1}^c} \mu_l^{k+1} \gamma_l^{k+1} < \frac{4L^2}{\bar{\rho}\sigma^r} \sum_{l \in \mathcal{B}_{k+1}^c} \mu_l^{k+1}, \quad k_r \leq k < k_{r+1}. \quad (4.9)$$

因此, 结合 (4.8) 和 (4.9) 可得

$$\epsilon_{k+1} = \sum_{l \in \mathcal{B}_{k+1}^f} \lambda_l^{k+1} (f(y^{k+1}) - \beta_l^{k+1}) - \sum_{l \in \mathcal{B}_{k+1}^c} \mu_l^{k+1} \gamma_l^{k+1} \leq \frac{4L^2}{\bar{\rho}\sigma^r}, \quad k_r \leq k < k_{r+1}.$$

故若

$$r \geq \frac{\ln \frac{4L^2}{\bar{\rho}\epsilon}}{\ln \sigma},$$

则有 $\epsilon_{k+1} \leq \epsilon$. 因此,

$$t_{\bar{k}} \leq \left\lceil \frac{\ln \frac{4L^2}{\bar{\rho}\epsilon}}{\ln \sigma} \right\rceil.$$

证毕. $\square$

为证明结论 (b) 和 (c), 先论证以下引理. 根据引理 4.1, 在以下两个引理中, 不妨假设束集不重置.

引理 4.2 设当前稳定中心不变, 束集不重置, $x^{\bar{k}}$ 为当前有效点, p 和 q 为两个迭代指标, 且 $p > q \geq \bar{k}$. 设 $(z_p, d^p), w_p$ 和 $(z_q, d^q), w_q$ 分别是 $\text{QP}(\mathcal{B}_p)$ 和 $\text{QP}(\mathcal{B}_q)$ 的最优解和最优值, 则

(1)

$$w_p \geq w_q + \frac{1}{2}\bar{\rho}\|d^p - d^q\|^2; \quad (4.10)$$

(2) 当 $q \rightarrow \infty$ 时, $\|d^p - d^q\| \rightarrow 0, z_p - z_q \rightarrow 0$.

证明 (1) 根据引理条件可知, (z_p, d^p) 为子问题 $\text{QP}(\mathcal{B}_q)$ 的一个可行解, 故有

$$\beta_l^q - f(y^q) + (g_{f_{il}}^l)^T d^p \leq z_p, \quad l \in \mathcal{B}_q^f, \quad (4.11)$$

$$\gamma_l^q + (g_{c_{jl}}^l)^T d^p \leq z_p, \quad l \in \mathcal{B}_q^c. \quad (4.12)$$

在 (4.11) 两端乘上 λ_l^q , 在 (4.12) 两端乘上 μ_l^q , 再相加, 得

$$\sum_{l \in \mathcal{B}_q^f} \lambda_l^q (\beta_l^q - f(y^q)) + \sum_{l \in \mathcal{B}_q^c} \mu_l^q \gamma_l^q + \left(\sum_{l \in \mathcal{B}_q^f} \lambda_l^q g_{f_{il}}^l + \sum_{l \in \mathcal{B}_q^c} \mu_l^q g_{c_{jl}}^l \right)^T d^p \leq z_p,$$

即

$$-\epsilon_q + (g^q)^T d^p \leq z_p. \quad (4.13)$$

又根据 (3.9), 有

$$\rho_q d^q = -g^q, \quad (4.14)$$

$$-\epsilon_q + (g^q)^T d^q = z_q. \quad (4.15)$$

结合 (4.13) 和 (4.15) 可知,

$$(g^q)^T (d^p - d^q) \leq z_p - z_q. \quad (4.16)$$

利用 (4.14) 和 (4.16) 以及 $\rho_p = \rho_q = \bar{\rho}$ 得

$$\begin{aligned} w_p &= z_p + \frac{1}{2}\bar{\rho}\|d^p\|^2 = z_p + z_q - z_q + \frac{1}{2}\bar{\rho}\|d^p + d^q - d^q\|^2 \\ &= w_q + \frac{1}{2}\bar{\rho}\|d^p - d^q\|^2 + z_p - z_q + \bar{\rho}(d^q)^T (d^p - d^q) \\ &= w_q + \frac{1}{2}\bar{\rho}\|d^p - d^q\|^2 + z_p - z_q - (g^q)^T (d^p - d^q) \\ &\geq w_q + \frac{1}{2}\bar{\rho}\|d^p - d^q\|^2. \end{aligned}$$

(2) 由 (4.10) 及 $w_k \leq 0$ 可知, 序列 $\{w_k\}_{k \geq \bar{k}}$ 单调非减且有上界, 故整列收敛, 于是再结合 (4.10) 可得

$$\|d^p - d^q\| \rightarrow 0, \quad q \rightarrow \infty.$$

上式结合假设 4.2 进一步推知,

$$\|d^p\|^2 - \|d^q\|^2 \rightarrow 0, \quad q \rightarrow \infty. \quad (4.17)$$

又由于

$$w_p - w_q = \frac{1}{2} \bar{\rho} (\|d^p\|^2 - \|d^q\|^2) + z_p - z_q,$$

故结合 (4.17) 有

$$z_p - z_q \rightarrow 0, \quad q \rightarrow \infty.$$

证毕. □

基于引理 4.2, 下面证明结论 (b) 和 (c) 成立.

引理 4.3 当稳定中心不变时, 算法只能有限次进入第 4 步, 且第 1 与第 3 步之间的循环也是有限的.

证明 反证法. 假设算法出现以下两种情形之一:

- (1) 无限次地进入第 4 步;
- (2) 在第 1 和第 3 步之间无限次地循环.

设 $x^{\bar{k}}$ 为当前有效点. 注意到 (1) 和 (2) 两种情形均意味着

$$\|g^k\| > \eta, \quad k \geq \bar{k} \quad (4.18)$$

成立无限次, 但不同时成立

$$\begin{aligned} f_i(x^{k+1}) - f(y^k) &\leq z_k + \delta_k, \quad i \in I, \\ c_j(x^{k+1}) &\leq 0, \quad j \in J. \end{aligned}$$

首先, 假设存在无限指标集 K , 使得

$$f_{i_k}(x^{k+1}) - f(y^k) > z_k + \delta_k, \quad k \in K,$$

其中, i_k 由第 3 步产生, 对应于指标 h .

设任意 $p, q \in K$, 且 $p > q$, 则

$$\begin{aligned} z_p &\geq (g_{f_{i_q}}^{q+1})^T d^p + \beta_{q+1}^{q+1} - f(y^{q+1}) \\ &= (g_{f_{i_q}}^{q+1})^T d^p + f_{i_q}(x^{q+1}) - (g_{f_{i_q}}^{q+1})^T d^q - f(y^{q+1}) \\ &> (g_{f_{i_q}}^{q+1})(d^p - d^q) + z_q + \delta_q. \end{aligned}$$

由于束集不重置, 故 $\delta_q \equiv \bar{\delta}$, 因此,

$$z_p - z_q > (g_{f_{i_q}}^{q+1})^T (d^p - d^q) + \bar{\delta},$$

两边对 $q \in K, q \rightarrow \infty$ 取极限, 再利用引理 4.2 和 (4.1) 得

$$\bar{\delta} \leq 0,$$

这与 $\bar{\delta} > 0$ 矛盾.

在下面分析中, 不妨假设存在 $j_k \in J$, 使得

$$c_{j_k}(x^{k+1}) > 0, \quad k \geq \bar{k}.$$

设指标 $p > q \geq \bar{k}$, 由于束集不重置, 故 $\rho_p = \rho_q = \bar{\rho}$, 从而有

$$\begin{aligned} 0 &\geq -\frac{1}{2}\bar{\rho}\|d^p\|^2 \geq z_p \geq (g_{c_{j_q}}^{q+1})^T d^p + \gamma_{q+1}^{q+1} = (g_{c_{j_q}}^{q+1})^T d^p + c_{j_q}(x^{q+1}) - (g_{c_{j_q}}^{q+1})^T d^q \\ &> (g_{c_{j_q}}^{q+1})^T (d^p - d^q), \end{aligned}$$

结合引理 4.2 和 (4.1) 得

$$\|d^p\| \rightarrow 0, \quad q \rightarrow \infty,$$

进而根据 (3.9) 有

$$\|g^p\| \rightarrow 0, \quad q \rightarrow \infty.$$

这与 (4.18) 矛盾, 故引理得证. □

由引理 4.1 和 4.3 立知下面定理成立.

定理 4.1 算法 3.1 是适定的, 即每经过有限次迭代之后, 必会更新稳定中心.

下面证明本文的主要结论: 算法 3.1 的全局收敛性.

定理 4.2 对任意 $\epsilon > 0$ 和 $\eta > 0$, 算法 3.1 经过有限多次迭代后终止, 并满足近似最优性条件

$$\|g^{k_*}\| \leq \eta, \quad \epsilon_{k_*} \leq \epsilon,$$

其中, k_* 为终止时的迭代指标. 此时, 若问题 (1.1) 满足 Slater 约束规格, 则 $y^{k_*} \in X$ 可作为其近似最优解.

证明 反证法. 假设算法 3.1 不能有限次终止, 则由定理 4.1 可知, 算法无限次地满足可行下降性, 并更新稳定中心.

记 k_s ($s = 1, 2, \dots$) 为有效迭代点的指标, 由引理 3.4 可知,

$$f(y^{k_{s+1}}) - f(y^{k_s}) \leq -\frac{1}{\rho_{k_{s+1}-1}} \|g^{k_{s+1}-1}\|^2 + \delta_{k_{s+1}-1}. \quad (4.19)$$

此外, 由算法第 1 步有

$$\|g^{k_{s+1}-1}\| > \eta. \quad (4.20)$$

设 t_{k_s} 是在由 y^{k_s} 变到 $y^{k_{s+1}}$ 的过程中束集重置的次数, 则

$$\rho_{k_{s+1}-1} = \bar{\rho} \sigma^{t_{k_s}}, \quad \delta_{k_{s+1}-1} = \frac{\bar{\delta}}{\sigma^{t_{k_s}}}. \quad (4.21)$$

结合 (4.19)–(4.21) 可得

$$f(y^{k_{s+1}}) - f(y^{k_s}) \leq \frac{1}{\sigma^{t_{k_s}}} \left(-\frac{\eta^2}{\bar{\rho}} + \bar{\delta} \right),$$

进而有

$$f(y^{k_{s+1}}) - f(y^{k_1}) \leq -\theta \sum_{l=1}^s \frac{1}{\sigma^{t_{k_l}}}, \quad (4.22)$$

其中,

$$\theta = \frac{\eta^2}{\rho} - \bar{\delta} > 0.$$

另一方面, 根据引理 4.1 可知,

$$t_{k_i} \leq \bar{t} := \left\lceil \frac{\ln \frac{4L^2}{\rho\epsilon}}{\ln \sigma} \right\rceil.$$

此关系结合 (4.22) 得

$$f(y^{k_{s+1}}) - f(y^{k_1}) \leq -\theta \sum_{l=1}^s \frac{1}{\sigma^{\bar{t}}}.$$

上式两边对 $s \rightarrow \infty$ 取极限, 得

$$\lim_{s \rightarrow \infty} (f(y^{k_{s+1}}) - f(y^{k_1})) \leq -\infty,$$

这与假设 4.1 矛盾. 证毕. $\square$

5 算法改进

随着算法迭代增加, 束集中的元素可能会无限增加, 从而导致数值计算和存储上的困难. 为此, 需要采取适当的措施, 既能控制束集中元素的数量在一个合理的范围, 又不破坏算法的收敛性. 目前主要有两种技术^[21]: 次梯度聚集和次梯度选取, 且使用前者居多. 因此, 本节将利用次梯度聚集技术, 对算法 3.1 进行改进.

5.1 次梯度聚集技术

为了更好地突出次梯度聚集技术的精髓, 本小节针对第 2.3 小节中的无约束优化束方法, 引入次梯度聚集技术. 随着迭代指标 k 的增大, 由 (2.6) 给出的束集 $\mathcal{B}_k$ 中的元素会不断累积, 进而子问题 (2.10) 的规模不断变大, 从而导致求解困难. 下面具体分析如何克服此困难.

设 (v_k, d^k) 为 (2.10) 的最优解, u^k 为相应的 KKT 乘子, 记聚集次梯度

$$(P_f^k, \tilde{\alpha}_P^k) = \sum_{l \in \mathcal{B}_k} u_l^k (g_f^l, \alpha_l^k), \quad (5.1)$$

则容易验证 (v_k, d^k) 亦为如下问题的唯一最优解 (详见文献 [21, 引理 2.2.2]),

$$\begin{aligned} \min_{v \in \mathbb{R}, d \in \mathbb{R}^n} \quad & v + \frac{1}{2} \rho_k \|d\|^2 \\ \text{s.t.} \quad & \alpha_l^k + (g_f^l)^T d \leq v, \quad l \in \tilde{\mathcal{B}}_k, \\ & \tilde{\alpha}_P^k + (P_f^k)^T d \leq v, \end{aligned} \quad (5.2)$$

其中, $\tilde{\mathcal{B}}_k$ 为 $\mathcal{B}_k$ 的任意子集, 甚至可以是空集. 下面分析如何根据 (5.2) 导出在新的迭代点 x^{k+1} 处的子问题.

首先, 结合 (2.3) 和 (2.9), 可将线性化函数表达为

$$\bar{f}^l(x) = \alpha_l^k + (g_f^l)^T (x - y^k), \quad l \in \mathcal{B}_k,$$

进而根据 (5.1) 可定义聚集的线性化函数

$$\tilde{f}^k(x) = \tilde{\alpha}_P^k + (P_f^k)^T(x - y^k),$$

及其在新的稳定中心 y^{k+1} 处的值

$$\alpha_P^{k+1} = \tilde{f}^k(y^{k+1}) = \tilde{\alpha}_P^k + (P_f^k)^T(y^{k+1} - y^k).$$

其次, 在新的稳定中心 y^{k+1} 处, 需要更新 α_l^k :

$$\alpha_l^{k+1} = f(x^l) + (g_f^l)^T(y^{k+1} - x^l) = \alpha_l^k + (g_f^l)^T(y^{k+1} - y^k), \quad l \in \tilde{\mathcal{B}}_k,$$

并且在 x^{k+1} 处产生新的线性化函数

$$\bar{f}^{k+1}(x) = f(x^{k+1}) + (g_f^{k+1})^T(x - x^{k+1}) = \alpha_{k+1}^{k+1} + (g_f^{k+1})^T(x - y^{k+1}),$$

其中,

$$\alpha_{k+1}^{k+1} = f(x^{k+1}) + (g_f^{k+1})^T(y^{k+1} - x^{k+1}), \quad g_f^{k+1} \in \partial f(x^{k+1}).$$

对应地, 相当于在子问题中加入新的约束

$$\alpha_{k+1}^{k+1} + (g_f^{k+1})^T d \leq v.$$

最后, 更新束集

$$\mathcal{B}_{k+1} = \tilde{\mathcal{B}}_k \cup \{(x^{k+1}, f(x^{k+1}), g_f^{k+1})\}.$$

综上分析, 在 x^{k+1} 处的 QP 子问题可写为

$$\begin{aligned} \min_{v \in \mathbb{R}, d \in \mathbb{R}^n} \quad & v + \frac{1}{2} \rho_{k+1} \|d\|^2 \\ \text{s.t.} \quad & \alpha_l^{k+1} + (g_f^l)^T d \leq v, \quad l \in \mathcal{B}_{k+1}, \\ & \alpha_P^{k+1} + (P_f^k)^T d \leq v. \end{aligned} \quad (5.3)$$

只要控制 $\tilde{\mathcal{B}}_k$ 的元素个数在合理的范围, 子问题 (5.3) 即可有效求解. 理论上可取 $\tilde{\mathcal{B}}_k = \emptyset$, 但为了获得较好的数值效果, 通常保留 $\mathcal{B}_k$ 中的部分元素.

5.2 带次梯度聚集的可行下降束方法

从上小节可看出, 次梯度聚集本质上就是对过往的次梯度作一个凸组合, 综合成一个“次梯度”, 进而将过往的线性化函数 (子问题约束) 聚集成一个函数 (约束).

由于子问题 (3.7) 包含有目标函数和约束函数的线性化, 因此需要分别对它们进行聚集. 结合子问题 (5.3), 并考虑递归求解, 在当前迭代指标 k 处, 求解如下聚集子问题:

$$\begin{aligned} \min_{z \in \mathbb{R}, d \in \mathbb{R}^n} \quad & z + \frac{1}{2} \rho_k \|d\|^2 \\ \text{AQP}(\mathcal{B}_k) \quad \text{s.t.} \quad & \beta_l^k - f(y^k) + (g_{f_{i_l}}^l)^T d \leq z, \quad l \in \mathcal{B}_k^f, \\ & \beta_P^k - f(y^k) + (P_f^{k-1})^T d \leq z, \\ & \gamma_l^k + (g_{c_{j_l}}^l)^T d \leq z, \quad l \in \mathcal{B}_k^c, \\ & \gamma_P^k + (P_c^{k-1})^T d \leq z, \end{aligned} \quad (5.4)$$

其中, $\mathcal{B}_k^f$ 和 $\mathcal{B}_k^c$ 在记号上虽然与子问题 (3.7) 相同, 但含义不同, 它们仅包含少量元素 (详见下面算法).

设 (z_k, d^k) 为 $\text{AQP}(\mathcal{B}_k)$ 的最优解, $(\lambda^k, \lambda_P^k) \in \mathbb{R}^{|\mathcal{B}_k^f|+1}$, $(\mu^k, \mu_P^k) \in \mathbb{R}^{|\mathcal{B}_k^c|+1}$ 为相应的 KKT 乘子, 则用于产生新子问题 $\text{AQP}(\mathcal{B}_{k+1})$ 的聚集次梯度更新如下:

$$(P_f^k, \tilde{\beta}_P^k) = \sum_{l \in \mathcal{B}_k^f} \tilde{\lambda}_l^k (g_{f_{i_l}}^l, \beta_l^k) + \tilde{\lambda}_P^k (P_f^{k-1}, \beta_P^k), \quad (5.5)$$

$$(P_c^k, \tilde{\gamma}_P^k) = \sum_{l \in \mathcal{B}_k^c} \tilde{\mu}_l^k (g_{c_{j_l}}^l, \gamma_l^k) + \tilde{\mu}_P^k (P_c^{k-1}, \gamma_P^k), \quad (5.6)$$

其中, $\tilde{\lambda}_l^k, l \in \mathcal{B}_k^f, \tilde{\lambda}_P^k, \tilde{\mu}_l^k, l \in \mathcal{B}_k^c, \tilde{\mu}_P^k$ 满足

$$\begin{aligned} \tau_f^k &= \sum_{l \in \mathcal{B}_k^f} \lambda_l^k + \lambda_P^k, \quad \lambda_l^k = \tau_f^k \tilde{\lambda}_l^k, \quad l \in \mathcal{B}_k^f, \quad \lambda_P^k = \tau_f^k \tilde{\lambda}_P^k, \\ \tilde{\lambda}_l^k &\geq 0, \quad l \in \mathcal{B}_k^f, \tilde{\lambda}_P^k \geq 0, \quad \sum_{l \in \mathcal{B}_k^f} \tilde{\lambda}_l^k + \tilde{\lambda}_P^k = 1, \\ \tau_c^k &= \sum_{l \in \mathcal{B}_k^c} \mu_l^k + \mu_P^k, \quad \mu_l^k = \tau_c^k \tilde{\mu}_l^k, \quad l \in \mathcal{B}_k^c, \quad \mu_P^k = \tau_c^k \tilde{\mu}_P^k, \\ \tilde{\mu}_l^k &\geq 0, \quad l \in \mathcal{B}_k^c, \quad \tilde{\mu}_P^k \geq 0, \quad \sum_{l \in \mathcal{B}_k^c} \tilde{\mu}_l^k + \tilde{\mu}_P^k = 1. \end{aligned}$$

易知,

$$\tau_f^k \geq 0, \quad \tau_c^k \geq 0, \quad \tau_f^k + \tau_c^k = 1.$$

此外, 选取束集 $\tilde{\mathcal{B}}_k^f \subseteq \mathcal{B}_k^f, \tilde{\mathcal{B}}_k^c \subseteq \mathcal{B}_k^c$, 并视具体情形产生 $\mathcal{B}_{k+1}^f$ 和 $\mathcal{B}_{k+1}^c$, 详见如下算法.

算法 5.1 (带次梯度聚集的可行下降束方法)

第 0 步 选取初始点 $x^1 \in X$, 令 $y^1 = x^1$, 根据 (3.14) 初始化束集, 并令

$$\beta_P^1 = \beta_1^1, \quad P_f^0 = g_{f_{i_1}}^1, \quad \gamma_P^1 = \gamma_1^1, \quad P_c^0 = g_{c_{j_1}}^1.$$

选取参数 $\eta > 0, \rho_1 = \bar{\rho} > 0$ 和 $\delta_1 = \bar{\delta} > 0$ 满足 $\eta^2 > \bar{\rho}\bar{\delta}$, 以及 $\epsilon > 0, \sigma > 1$, 令 $W = I \cup J, k = 1$.

第 1 步 求解子问题 $\text{AQP}(\mathcal{B}_k)$ 得最优解 (z_k, d^k) 以及相应的乘子 $(\lambda^k, \lambda_P^k, \mu^k, \mu_P^k)$, 并根据 (5.5) 和 (5.6) 产生 $(P_f^k, \tilde{\beta}_P^k)$ 和 $(P_c^k, \tilde{\gamma}_P^k)$. 选取束集

$$\tilde{\mathcal{B}}_k^f \subseteq \mathcal{B}_k^f, \quad \tilde{\mathcal{B}}_k^c \subseteq \mathcal{B}_k^c.$$

令 $x^{k+1} = y^k + d^k$, 计算

$$g^k = \sum_{l \in \mathcal{B}_k^f} \lambda_l^k g_{f_{i_l}}^l + \lambda_P^k P_f^{k-1} + \sum_{l \in \mathcal{B}_k^c} \mu_l^k g_{c_{j_l}}^l + \mu_P^k P_c^{k-1}.$$

若 $\|g^k\| > \eta$, 进入第 3 步.

第 2 步 计算

$$\epsilon_k = \sum_{l \in \mathcal{B}_k^f} \lambda_l^k (f(y^k) - \beta_l^k) + \lambda_P^k (f(y^k) - \beta_P^k) - \sum_{l \in \mathcal{B}_k^c} \mu_l^k \gamma_l^k - \mu_P^k \gamma_P^k.$$

若 $\epsilon_k \leq \epsilon$, 则算法终止 (满足近似最优性条件); 否则, 令 $y^{k+1} = y^k$, 由 (3.15) 和 (3.16) 重置束集, 并令

$$\begin{aligned} \beta_P^{k+1} &= \beta_{l(k+1)}^{k+1}, \quad P_f^k = g_f^{l(k+1)}, \\ \begin{cases} \gamma_P^{k+1} = \gamma_{l(k+1)}^{k+1}, & P_c^k = g_c^{l(k+1)}, & \text{若 } c(x^{l(k+1)}) = 0, \\ \gamma_P^{k+1} = -\infty, & P_c^k = 0, & \text{若 } c(x^{l(k+1)}) < 0. \end{cases} \end{aligned}$$

令

$$\rho_{k+1} = \sigma \rho_k, \quad \delta_{k+1} = \frac{\delta_k}{\sigma}, \quad k := k + 1,$$

返回第 1 步.

第 3 步 选取一个指标 $h \in W$.

(1) 若 $h \in I$, 且 $f_h(x^{k+1}) - f(y^k) > z_k + \delta_k$, 则令 $y^{k+1} = y^k$ (无效步), 并令

$$\begin{aligned} \mathcal{B}_{k+1}^f &= \tilde{\mathcal{B}}_k^f \cup \{(x^{k+1}, f_h(x^{k+1}), g_{f_h}^{k+1}, \beta_{k+1}^{k+1})\}, \quad \mathcal{B}_{k+1}^c = \tilde{\mathcal{B}}_k^c, \\ \beta_P^{k+1} &= \tilde{\beta}_P^k, \quad \gamma_P^{k+1} = \tilde{\gamma}_P^k. \end{aligned}$$

恢复指标集 $W = I \cup J$, 令 $\rho_{k+1} = \rho_k$, $\delta_{k+1} = \delta_k$, $k := k + 1$, 返回第 1 步.

(2) 若 $h \in J$, 且 $c_h(x^{k+1}) > 0$, 则令 $y^{k+1} = y^k$ (无效步), 并令

$$\begin{aligned} \mathcal{B}_{k+1}^c &= \tilde{\mathcal{B}}_k^c \cup \{(x^{k+1}, c_h(x^{k+1}), g_{c_h}^{k+1}, \gamma_{k+1}^{k+1})\}, \quad \mathcal{B}_{k+1}^f = \tilde{\mathcal{B}}_k^f, \\ \beta_P^{k+1} &= \tilde{\beta}_P^k, \quad \gamma_P^{k+1} = \tilde{\gamma}_P^k. \end{aligned}$$

恢复指标集 $W = I \cup J$, 令 $\rho_{k+1} = \rho_k$, $\delta_{k+1} = \delta_k$, $k := k + 1$, 返回第 1 步.

第 4 步 令 $W := W \setminus \{h\}$. 若 $W \neq \emptyset$, 返回第 1 步.

第 5 步 令 $y^{k+1} = x^{k+1}$ (有效步, 更新稳定中心, x^{k+1} 为新的有效点), 分别由 (3.19) 和 (3.20) 产生 β_l^{k+1} , $l \in \tilde{\mathcal{B}}_k^f$ 和 γ_l^{k+1} , $l \in \tilde{\mathcal{B}}_k^c$, 并令

$$\begin{aligned} \mathcal{B}_{k+1}^f &= \{(x^l, f_{i_l}(x^l), g_{f_{i_l}}^l, \beta_l^{k+1}), l \in \tilde{\mathcal{B}}_k^f\} \cup \{(x^{k+1}, f(x^{k+1}), g_f^{k+1}, \beta_{k+1}^{k+1})\}, \\ \mathcal{B}_{k+1}^c &= \{(x^l, c_{j_l}(x^l), g_{c_{j_l}}^l, \gamma_l^{k+1}), l \in \tilde{\mathcal{B}}_k^c\} \cup \{(x^{k+1}, c(x^{k+1}), g_c^{k+1}, \gamma_{k+1}^{k+1})\}, \\ \beta_P^{k+1} &= \tilde{\beta}_P^k + (P_f^k)^T(y^{k+1} - y^k), \quad \gamma_P^{k+1} = \tilde{\gamma}_P^k + (P_c^k)^T(y^{k+1} - y^k). \end{aligned}$$

恢复指标集 $W = I \cup J$, 令 $\rho_{k+1} = \bar{\rho}$, $\delta_{k+1} = \bar{\delta}$, $k := k + 1$, 返回第 1 步.

注 5.1 由于引入次梯度聚集之后, 不会影响算法的收敛性, 因此, 算法 5.1 和 3.1 具有同样的收敛性质.

6 数值试验

为了验证算法 5.1 的实际有效性, 本节利用 MATLAB 编程, 进行了初步的数值试验. 测试的 8 个问题构成如下: 问题 P1–P4 分别为文献 [13] 的问题 5–7 和 2; 问题 P5 和 P6 的目标函数为文献 [35] 的问题 Max1, 约束函数分别为文献 [36] 的问题 4.6(1) 和 4.6(2); 问题 P7 和 P8 的目标函数为文献 [36] 的问题 2.1, 约束函数分别为文献 [36] 的问题 4.6(1) 和 4.6(2). 这些问题当中, 有些分量函数是光滑的,

有些是非光滑的. 当分量函数光滑时 (次微分是单点集, 即 $\partial f(x) = \{\nabla f(x)\}$), 我们直接取其梯度; 当分量函数非光滑时, 将根据函数的特殊结构产生其次梯度, 例如, 对函数 $f_i(x) = |x_i|$, 取其次梯度为

$$g_{f_i} = \begin{cases} e_i, & x_i \geq 0, \\ -e_i, & x_i < 0, \end{cases}$$

其中, $e_i \in \mathbb{R}^n$ 表示第 i 个元素为 1、其余元素为 0 的向量.

算法参数选取为 $\eta = 0.02$, $\bar{\rho} = 0.7$, $\bar{\delta} = 0.00055$, $\epsilon = 0.001$, $\sigma = 10$, 其中, 终止参数 η 和 ϵ 的取值很大程度上决定了算法计算结果的精确性, 而经过观察, 算法对参数 $\bar{\rho}$, $\bar{\delta}$ 和 σ 的取值相对比较稳定, 即当参数发生微小改变时, 算法不会产生较大的波动.

在第 1 步中, 束集 $\tilde{\mathcal{B}}_k^f$ 和 $\tilde{\mathcal{B}}_k^c$ 产生如下: 设定一个束集最大容限 N (计算时取 $N = 10n$), 若 $|\mathcal{B}_k^f| \leq N$, 则取 $\tilde{\mathcal{B}}_k^f = \mathcal{B}_k^f$, 否则取 $\mathcal{B}_k^f$ 中最新产生的 N 个元素组成 $\tilde{\mathcal{B}}_k^f$. 同样的方法应用于产生 $\tilde{\mathcal{B}}_k^c$.

在第 1 步中, 我们采用一种较为简单的策略选取指标 h : 在集合 $W = \{1, \dots, m+p\}$ 中顺序选取的基础上, 优先选取那些以往违背目标下降性或约束可行性的指标.

数值试验结果列示在表 1 中, 其中部分符号含义如下, **ni**: 算法总迭代次数; **nsi**: 算法有效迭代次数; **nf**: 目标分量函数的计算次数; **nc**: 约束分量函数的计算次数; **nfg**: 目标分量函数次梯度的计算次数; **ncg**: 约束分量函数次梯度的计算次数; f^* : 算法终止时的近似最优值.

从表 1 可以看出, 对给出的 8 个测试问题, 本文算法均能快速、有效地求解. 当然, 需要指出的是, 由于目前公开的求解非光滑约束优化的软件极少, 我们还没有跟其他方法作数值比较, 但将在未来工作中继续完善.

表 1 算法 5.1 的数值试验报告

问题	n, m, p	x^1	ni(nsi)	nf	nc	nfg	ncg	f^*
P1	4, 4, 3	$(0, 1, 1, 0)^T$	85(13)	141	39	83	13	-4.399993e+01
P2	2, 3, 2	$(0, 1)^T$	41(6)	70	12	39	6	2.000013e+00
P3	7, 5, 4	$(1, 2, 0, 4, 0, 1, 1)^T$	375(11)	489	44	372	11	6.806302e+02
P4	2, 6, 2	$(-1, 1)^T$	17(5)	51	10	17	5	6.235882e-01
P5	10, 10, 9	$(1, \dots, 1)^T$	61(17)	409	155	55	19	3.333332e-01
	20, 20, 19	$(1, \dots, 1)^T$	100(16)	1181	305	95	17	3.333327e-01
	50, 50, 49	$(1, \dots, 1)^T$	231(31)	6617	1519	228	31	3.333369e-01
P6	10, 10, 9	$(0.5, \dots, 0.5)^T$	15(5)	105	45	15	5	6.938894e-17
	20, 20, 19	$(0.5, \dots, 0.5)^T$	28(8)	370	152	28	8	5.551115e-17
	50, 50, 49	$(0.5, \dots, 0.5)^T$	69(19)	2225	931	69	19	2.081668e-17
P7	10, 10, 9	$(0.1, \dots, 0.1)^T$	35(6)	192	54	35	6	7.019094e-04
	20, 20, 19	$(0.1, \dots, 0.1)^T$	64(9)	745	171	64	9	1.030068e-03
	50, 50, 49	$(0.1, \dots, 0.1)^T$	111(11)	3005	539	111	11	2.866618e-03
P8	10, 10, 9	$(0.5, \dots, 0.5)^T$	111(14)	537	128	105	16	1.111111e-01
	20, 20, 19	$(0.5, \dots, 0.5)^T$	215(12)	2007	229	210	13	1.111116e-01
	50, 50, 49	$(0.5, \dots, 0.5)^T$	476(16)	11003	784	473	16	1.111140e-01

7 结语

本文针对不等式约束 Minimax 问题的特殊结构, 提出了一个可行下降束方法. 该方法属于非光滑方法, 用函数的次梯度取代光滑情形下的梯度来设计算法. 通过利用部分割平面技术, 避免计算所有分量函数的函数值, 从而有效减少了计算量. 算法不仅能产生可行的有效迭代点, 而且能保证在有效点列上目标函数值单调下降. 通过次梯度聚集技术对束集中的元素进行聚集, 提升了算法的实际执行效率. 此外, 证明了算法的全局收敛性, 且初步的数值试验说明算法是有效的.

作为本文方法的进一步工作, 还有一些问题值得继续深入思考和研究. 例如, 采用罚函数法、两阶段方法或强次可行方法等 (参见文献 [21, 22, 37]), 推广到初始点不可行的情形; 考虑分量函数为非凸的情形, 即假设分量函数 Lipschitz 连续, 用 Clarke 次梯度代替凸情形下的次梯度, 但此时的部分割平面模型将不再构成下近似模型, 从而带来不小的困难. 近年来, Kiwiel, Lemaréchal, Sagastizábal 等学者对基于非精确数据的束方法进行了研究, 并成功应用于求解一些实际问题, 如文献 [38–40]. 因此, 可考虑将本文的方法与这类非精确数据方法结合, 得到一个基于非精确数据的版本; 本文从定性的角度证明了算法的全局收敛性, 今后可考虑从定量的角度进一步分析算法的复杂度.

致谢 感谢两位审稿人提出的宝贵的建设性意见.

参考文献

- 1 Michelot C, Plastria F. An extended multifacility minimax location problem revisited. *Ann Oper Res*, 2002, 111: 167–179
- 2 Wang S Y, Yamamoto Y, Yu M. A minimax rule for portfolio selection in frictional markets. *Math Methods Oper Res*, 2003, 57: 141–155
- 3 Chernousko F L. Minimax control for a class of linear systems subject to disturbances. *J Optim Theory Appl*, 2005, 127: 535–548
- 4 Baums A. Minimax method in optimizing energy consumption in real-time embedded systems. *Autom Control Comput Sci*, 2009, 43: 57–62
- 5 Li Y P, Huang G H. Inexact minimax regret integer programming for long-term planning of municipal solid waste management — Part A: Methodology development. *Environ Eng Sci*, 2009, 26: 209–218
- 6 Watson G A. The minimax solution of an overdetermined system of nonlinear equations. *IMA J Appl Math*, 1979, 23: 167–180
- 7 Overton M L. Algorithms for nonlinear ℓ_1 and ℓ_∞ fitting. In: *Nonlinear Optimization*. London: Academic Press, 1982, 213–221
- 8 Li X S, Fang S C. On the entropic regularization method for solving min-max problems with applications. *Math Methods Oper Res*, 1997, 46: 119–130
- 9 Xu S. Smoothing method for minimax problems. *Comput Optim Appl*, 2001, 20: 267–279
- 10 Di Pillo G, Grippo L, Lucidi S. A smooth method for the finite minimax problem. *Math Program*, 1993, 60: 187–214
- 11 Vardi A. New minimax algorithm. *J Optim Theory Appl*, 1992, 75: 613–634
- 12 Wang F S, Zhang K C. A hybrid algorithm for nonlinear minimax problems. *Ann Oper Res*, 2008, 164: 167–191
- 13 Rustem B, Nguyen Q. An algorithm for the inequality constrained discrete minimax problem. *SIAM J Optim*, 1998, 8: 265–283
- 14 Jian J B, Zhang X L, Quan R, et al. Generalized monotone line search SQP algorithm for constrained minimax problems. *Optimization*, 2009, 58: 101–131
- 15 Jian J B, Mo X D, Qiu L J, et al. Simple sequential quadratically constrained quadratic programming feasible algorithm with active identification sets for constrained minimax problems. *J Optim Theory Appl*, 2014, 160: 158–188
- 16 Gaudioso M, Giallombardo G, Miglionico G. An incremental method for solving convex finite min-max problems. *Math Oper Res*, 2006, 31: 173–187
- 17 Gaudioso M, Giallombardo G, Miglionico G. On solving the lagrangian dual of integer programs via an incremental approach. *Comput Optim Appl*, 2009, 44: 117–138

- 18 Fuduli A, Gaudioso M, Giallombardo G, et al. A partially inexact bundle method for convex semi-infinite minmax problems. *Commun Nonlinear Sci Numer Simul*, 2015, 21: 172–180
- 19 Hare W, Macklem M. Derivative-free optimization methods for finite minimax problems. *Optim Methods Softw*, 2013, 28: 300–312
- 20 Hare W, Nutini J. A derivative-free approximate gradient sampling algorithm for finite minimax problems. *Comput Optim Appl*, 2013, 56: 1–38
- 21 Kiwiel K C. *Methods of Descent for Nondifferentiable Optimization*. Berlin: Springer-Verlag, 1985
- 22 Bonnans J F, Gilbert J C, Lemaréchal C, et al. *Numerical Optimization: Theoretical and Practical Aspects*. 2nd ed. New York: Springer-Verlag, 2006
- 23 Burke J V, Lewis A S, Overton M L. A robust gradient sampling algorithm for nonsmooth, nonconvex optimization. *SIAM J Optim*, 2005, 15: 751–779
- 24 Kiwiel K C. Exact penalty functions in proximal bundle methods for constrained convex nondifferentiable minimization. *Math Program*, 1991, 52: 285–302
- 25 Curtis F E, Overton M L. A sequential quadratic programming algorithm for nonconvex, nonsmooth constrained optimization. *SIAM J Optim*, 2012, 22: 474–500
- 26 Sagastizábal C, Solodov M. An infeasible bundle method for nonsmooth convex constrained optimization without a penalty function or a filter. *SIAM J Optim*, 2005, 16: 146–169
- 27 Karas E, Ribeiro A, Sagastizábal C, et al. A bundle-filter method for nonsmooth convex constrained optimization. *Math Program*, 2009, 116: 297–320
- 28 Cheney E, Goldstein A. Newton's method for convex programming and Tchebycheff approximations. *Numer Math*, 1959, 1: 253–268
- 29 Kelley J E. The cutting plane method for solving convex programs. *J Soc Indust Appl Math*, 1960, 8: 703–712
- 30 Nedic A, Bertsekas D P. Incremental subgradient methods for nondifferentiable optimization. *SIAM J Optim*, 2001, 12: 109–138
- 31 Yuan Y X, Sun W Y. *Optimization Theory and Methods*. Beijing: Science Press, 1997
- 32 Kiwiel K C. Restricted step and Levenberg-Marquardt techniques in proximal bundle methods for nonconvex nondifferentiable optimization. *SIAM J Optim*, 1996, 6: 227–249
- 33 Jian J B. *Fast Algorithms for Smooth Constrained Optimization — Theoretical Analysis and Numerical Experiments*. Beijing: Science Press, 2010
- 34 Kiwiel K C. A method for solving certain quadratic programming problems arising in nonsmooth optimization. *IMA J Numer Anal*, 1986, 6: 137–152
- 35 Lukšan L, Vlček J. *Test Problems for Nonsmooth Unconstrained and Linearly Constrained Optimization*. Technical Report. Prague: Academy of Sciences of the Czech Republic, 2000
- 36 Karmitsa N. *Test Problems for Large-Scale Nonsmooth Minimization*. Technical Report. Jyväskylä: University of Jyväskylä, 2007
- 37 Tang C M, Jian J B. Strongly sub-feasible direction method for constrained optimization problems with nonsmooth objective functions. *European J Oper Res*, 2012, 218: 28–37
- 38 Kiwiel K C, Lemaréchal C. An inexact bundle variant suited to column generation. *Math Program*, 2007, 118: 177–206
- 39 de Oliveira W, Sagastizábal C, Lemaréchal C. Convex proximal bundle methods in depth: A unified analysis for inexact oracles. *Math Program*, 2014, 148: 241–277
- 40 van Ackooij W, Sagastizábal C. Constrained bundle methods for upper inexact oracles with application to joint chance constrained energy problems. *SIAM J Optim*, 2014, 24: 733–765

A feasible descent bundle method for inequality constrained Minimax problems

JIAN JinBao, TANG ChunMing & TANG Fei

Abstract In this paper, a feasible descent bundle method for solving inequality constrained Minimax problems is proposed. The main features of the method are: (1) By using the subgradients of functions and the idea of bundle method, it does not assume that the component functions of the original problem are smooth; (2) by adopting

the technique of partial cutting-planes model, at each null step, only the function value and subgradient of one component function are needed to generate the new cutting plane, and therefore the computational cost is reduced effectively; (3) it can ensure the feasibility of the serious iterates and the descent of the objective function; (4) by introducing the technique of subgradient aggregation to aggregate the subgradients in the bundle, the difficulty of numerical calculation and storage is overcome; (5) the algorithm is proved to be globally convergent, and the preliminary numerical results show that the proposed algorithm is effective.

Keywords inequality constraints, Minimax problems, feasible descent, bundle method, partial cutting-planes model

MSC(2010) 90C30, 65K05

doi: 10.1360/N012015-00022