

最小约束违背非线性凸优化

路斯文

长沙理工大学数学与统计学院, 湖南 长沙

收稿日期: 2024年8月10日; 录用日期: 2024年9月2日; 发布日期: 2024年9月12日

摘要

本文基于不可行性度量和互补约束优化模型的角度研究最小约束违背凸优化问题。首先我们对约束不相容的凸优化问题建立了最小约束违背优化模型。当问题中的约束相容时, 该模型可退化为原始问题。当约束不相容时, 该模型等价于某个MPCC问题。其次我们证明了该等价问题的W-稳定性。最后我们用增广拉格朗日方法求解该等价问题, 证明了该方法生成的点列收敛到等价MPCC问题的W-稳定点。

关键词

最小约束违背优化问题, MPCC, W-稳定点, 增广拉格朗日方法

Nonlinear Convex Optimization with Least Constraint Violation

Siwen Lu

School of Mathematics and Statistics, Changsha University of Science and Technology, Changsha Hunan

Received: Aug. 10th, 2024; accepted: Sep. 2nd, 2024; published: Sep. 12th, 2024

Abstract

In this paper, the problem of least constrained contracovex optimization is studied from the perspective of the infeasibility measure and the complementary constraint optimization model. Firstly, we establish a minimum constraint violation optimization model for the convex optimization problem with incompatible constraints. When the constraints in the problem are compatible, the model can degenerate to the original problem. When the constraints are incompatible, the model is equivalent to an MPCC problem. Second, we demonstrate the W-stability of the equivalence problem. Finally, we use the augmented Lagrangian method to solve the equivalence problem, and prove that the point series generated by the method converges to the W-stable point of the equivalent MPCC problem.

文章引用: 路斯文. 最小约束违背非线性凸优化[J]. 应用数学进展, 2024, 13(9): 4119-4128.

DOI: 10.12677/aam.2024.139393

Keywords

Least Constraint Violation, MPCC, W-Stationary Point, Augmented Lagrangian Method

Copyright © 2024 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

关于非线性规划问题, 已有不少不可行稳定点和算法。例如 Byrd, Curtis 与 Nocedal [1] 提出了 SQP 算法, 在一组条件下可以保证算法超线性收敛到一个 l_1 -范数度量的不可行稳定点。Burke, Curtis 与 Wang [2] 考虑具有等式和不等式约束的一般非线性规划问题, 证明了他们所提的 SQP 算法可以全局并且快速收敛到 KKT 点, 或者超线性或二次收敛到 l_1 -范数度量的不可行稳定点。Dai, Liu 与 Sun [3] 提出了一个原始-对偶内点方法, 并证明了该内点方法当原始问题可行时, 在一定条件下超线性或二次收敛到问题的 KKT 点; 当问题不可行时, 则超线性或二次收敛到 l_2 -范数度量的不可行稳定点。

这些算法都是求解约束条件不可行性度量的某个稳定点, 这样的稳定点与原问题的目标函数没有任何关系, 并不是最小约束违背优化问题的解[4]。因此有必要考虑在最小约束违背的点集上极小化目标函数。Dai 和 Zhang [5] 依据 Lipschitz 连续优化的最优性理论, 提出了 L-稳定点的概念, 并构造了求解最小约束违背的非线性凸规划问题的惩罚函数方法和用于求解最小约束违背的非线性凸优化问题等价 MPCC 问题的光滑 Fischer-Burmeister 函数方法。

然而, 惩罚函数方法并不是精确的, 当惩罚参数较大时, 惩罚函数方法会出现计算困难, 并且光滑函数方法只能收敛到问题的 L-稳定点。由于约束规范不同导致无法确定 L-稳定点和 S-稳定点之间的关系。为了克服惩罚函数方法和光滑函数方法在求解最小约束违背优化问题的局限性, 对最小约束违背凸优化问题值得重新考虑新的理论和算法。Chiche 和 Gilbert [6] 的研究为我们提供了有价值的线索, 他们证明了增广拉格朗日方法可以处理一个不可行的凸二次规划问题, 这促使我们考虑用增广拉格朗日方法来处理最小约束违背的非线性凸优化问题。

本文采用增广拉格朗日方法求解最小约束违背优化等价的 MPCC 问题, 进而达到求解最小约束违背优化问题的目的, 并证明了增广拉格朗日方法生成的点列收敛到该等价 MPCC 问题的 W-稳定点。

2. 基于不可行性度量的数学规划模型

本节考虑如下形式的非线性规划问题

$$\begin{aligned} \min & f(x) \\ \text{s.t. } & h(x) = 0, \\ & g(x) \geq 0, \end{aligned} \quad (2.1)$$

其中用 $f: R^n \rightarrow R$, $h: R^n \rightarrow R^q$, $g: R^n \rightarrow R^p$ 。记问题(2.1)的可行域:

$$\Phi = \{x \in R^n : h(x) = 0, g(x) \geq 0\}.$$

基于上述非线性规划问题, 为后续讨论需要, 引入问题的不可行性度量。

定义 2.1 [5] 函数 $\theta: R^n \rightarrow R$ 称为约束 $(h(x), g(x)) \in \{0_q\} \times R_+^p$ 的不可行度量, 如果存在递增连续函数 $\rho: R_+ \rightarrow R_+$ 满足 $\rho(0) = 0$, 使得

$$\theta(x) = \rho(\text{dist}((h(x), g(x)), \{0_q\} \times R_+^p)),$$

其中

$$\text{dist}((h(x), g(x)), \{0_q\} \times R_+^p) = \inf \{ \|h(x), y - g(x)\| : y \in R_+^p \}$$

是由 $(h(x), g(x))$ 到 $\{0_q\} \times R_+^p$ 在 $R^q \times R^p$ 的范数 $\|\cdot\|$ 下的距离。

显然, $\theta(x)$ 依赖于函数 $\rho(\cdot)$ 和范数 $\|\cdot\|$ 。本文中采用 $\rho(t) = \frac{1}{2}t^2$, 范数 $\|\cdot\|$ 采用标准的欧式范数, 即 l_2 -范数。

在上述不可行性度量下, 引入在具有最小约束违背点集上极小化目标函数 $f(x)$ 的数学模型。

定义 2.2 [5] 对于约束 $(h(x), g(x)) \in \{0_q\} \times R_+^p$ 的一个不可行度量 $\theta(x)$, 与 θ 相联系的最小约束违背的极小化目标函数 $f(x)$ 的数学模型定义为

$$\begin{aligned} \min & f(x) \\ \text{s.t. } & x \in \underset{z}{\text{Arg min}} \theta(z). \end{aligned} \quad (2.2)$$

显然, 如果可行域 Φ 非空, 那么 $\min_z \theta(z) = 0$, $\underset{z}{\text{Arg min}} \theta(z) = \Phi$, 问题(2.2)恰好就是原始问题(2.1)。因此, 问题(2.2)可以被视为原始问题(2.1)的拓广。

在定义 2.2 中, 没有阐述 $\underset{z}{\text{Arg min}} \theta(z)$ 的含义, 要看具体情况如何约定。如果 $\theta(z)$ 是凸函数, 显然这一集合就是全局极小点的集合。然而, 如果 $\theta(z)$ 是非凸函数(比如问题是可行的, 当 Φ 是非凸时就是这一情形)。 $\underset{z}{\text{Arg min}} \theta(z)$ 可以理解为局部极小点, 甚至是稳定点集。在非凸优化的情形, 不可行性检测是非常难的问题。

对 $\rho(t) = \frac{1}{2}t^2$, $t \geq 0$, 约束的最小违背定义为下述问题的最优值

$$\begin{aligned} \min & \frac{1}{2} \| (h(x), y) \|^2 \\ \text{s.t. } & g(x) + y \geq 0_p. \end{aligned} \quad (2.3)$$

最小约束违背的点集定义为

$$S = \{x : (x, y) \text{ 求解问题 (2.3)}\}.$$

在 S 上极小化 f 的问题是

$$\begin{aligned} \min & f(x) \\ \text{s.t. } & (x, y) \text{ 求解} \\ & \left\{ \begin{aligned} \min_{w, z} & \frac{1}{2} \|h(w)\|^2 + \|z\|^2 \\ \text{s.t. } & g(w) + z \geq 0. \end{aligned} \right. \end{aligned} \quad (2.4)$$

用 (P_L) 记问题(2.4)的下层问题, 即

$$(P_L) \quad \left\{ \begin{aligned} \min_{w, z} & \left[\frac{1}{2} \|h(w)\|^2 + \|z\|^2 \right] \\ \text{s.t. } & g(w) + z \geq 0. \end{aligned} \right.$$

利用与 $\rho(t)$ 的关系, 有

$$\theta(w) = \frac{1}{2} \|h(x)\|^2 + \min_z \left\{ \frac{1}{2} \|z\|^2 : g(w) + z \geq 0 \right\} = \frac{1}{2} \left[\|h(w)\|^2 + \|[g(w)]_-\|^2 \right],$$

其中 $[z]_- = ([z_1]_-, \dots, [z_p]_-)^T$, $z \in R^p$, 其中 $[t]_- = \min\{0, t\}$, $t \in R$ 。容易验证, $\theta(w)$ 是问题(2.1)的一个不可行性度量。问题 (P_L) 的最优解可表示为

$$\text{Arg min}(P_L) = \text{Arg min}_x \theta(x) = \text{Arg min}_x \frac{1}{2} \left[\|h(x)\|^2 + \|[g(x)]_-\|^2 \right].$$

所以, 问题(2.4)可等价地表示为

$$\begin{aligned} & \min f(x) \\ & \text{s.t. } x \in \text{Arg min}_w \frac{1}{2} \left[\|h(w)\|^2 + \|[g(w)]_-\|^2 \right]. \end{aligned} \quad (2.5)$$

容易验证, 如果 h 和 g 是可微的, 则 $\theta(x)$ 是可微的, 且满足

$$\nabla \theta = Jh(x)^T h(x) + Jg(x)^T [g(x)]_-. \quad (2.6)$$

以下讨论最小约束违背的非线性凸优化问题, 将从模型(2.5)出发, 并利用公式(2.6)解决该问题。

3. 最小约束违背非线性凸优化

3.1. 最小约束违背非线性凸优化等价的 MPCC 问题

考虑下述非线性凸优化问题

$$\begin{aligned} & \min f(x) \\ & \text{s.t. } Ax - b = 0, \\ & \quad g(x) \geq 0, \end{aligned} \quad (3.1)$$

其中 f 是一光滑的凸函数, $A \in R^{q \times n}$, $b \in R^q$, 且每一个 $g_i (i=1, \dots, p)$ 都是凹的光滑函数。此种情况下, 函数 $\theta(x)$ 是凸函数, 从而问题(2.5)是凸优化问题, 可被简化为

$$\begin{aligned} & \min f(x) \\ & \text{s.t. } A^T (Ax - b) + Jg(x)^T [g(x)]_- = 0. \end{aligned} \quad (3.2)$$

尽管问题(3.2)是一个凸优化问题, 其约束处理并不容易, 因为这些约束是非光滑的等式, 需要将约束转化为光滑约束, 构造数值算法。通过引入辅助向量 $y \in R^p$, 可将(3.2)中的约束表示为

$$\begin{aligned} & A^T (Ax - b) + Jg(x)^T y = 0, \\ & 0 \leq y \perp g(x) + y \geq 0. \end{aligned}$$

为讨论方便, 定义 $z = g(x) + y$, 上述系统可以表示为

$$F(x, y, z) = 0, (y, z) \in \Omega,$$

其中

$$\begin{aligned} F(x, y, z) &= \begin{pmatrix} A^T (Ax - b) + Jg(x)^T y \\ g(x) + y - z \end{pmatrix}, \\ \Omega &= \{(y, z) \in R^p \times R^q : 0 \leq y \perp z \geq 0\}. \end{aligned} \quad (3.3)$$

因此, 问题(2.5)等价地表示为

$$\begin{aligned} \min & f(x) \\ \text{s.t. } & F(x, y, z) = 0, (y, z) \in \Omega. \end{aligned} \quad (3.4)$$

映射 F 在 (x, y, z) 的 Jacobian 矩阵具有形式

$$JF(x, y, z) = \begin{pmatrix} A^T A - \sum_{j=1}^p y_j \nabla^2 g_j(x) & -Jg(x)^T & 0 \\ Jg(x) & I_p & -I_p \end{pmatrix}. \quad (3.5)$$

为后续讨论方便, 我们记 $H(x, y) = A^T A - \sum_{j=1}^p y_j \nabla^2 g_j(x)$, $G(x) = Jg(x)$ 。

用 Φ 记问题(3.4)的可行域, 即

$$\Phi = \{(x, y, z) \in R^n \times \Omega : F(x, y, z) = 0\}.$$

那么问题(3.4)可被简化为如下的 MPCC 问题

$$\begin{aligned} \min & f(x) \\ \text{s.t. } & (x, y, z) \in \Phi. \end{aligned} \quad (3.6)$$

3.2. 等价 MPCC 问题的必要性最优性条件

本节主要讨论 MPCC (3.6)的最优性条件, 下面我们首先来回顾一些符号和定理。定义下述指标集合

$$\begin{aligned} \alpha &:= \{i \mid y_i = 0, z_i > 0\}, \\ \beta &:= \{i \mid y_i = 0, z_i = 0\}, \\ \gamma &:= \{i \mid y_i > 0, z_i = 0\}. \end{aligned}$$

为了得到 $N_\Phi(\bar{x}, \bar{y}, \bar{z})$ 的显示表达式, 我们首先给出抽象约束集合的切锥、法锥形式。

定理 3.1 [7] 设 $\Omega_i = \{(y_i, z_i) \in R^2 \mid y_i \geq 0, z_i \geq 0, y_i z_i = 0\}$, 则对 $\forall (y_i, z_i) \in \Omega_i$, 有

$$\begin{aligned} T_{\Omega_i}(y_i, z_i) &= \begin{cases} \{0\} \times R, & y_i = 0, z_i > 0, \\ R \times \{0\}, & y_i > 0, z_i = 0, \\ \Omega_i, & y_i = 0, z_i = 0, \end{cases} \\ N_{\Omega_i}(y_i, z_i) &= \begin{cases} R \times \{0\}, & y_i = 0, z_i > 0, \\ \{0\} \times R, & y_i > 0, z_i = 0, \\ (R_- \times R_-) \cup (\{0\} \times R) \cup (R \times \{0\}), & y_i = 0, z_i = 0, \end{cases} \\ \hat{N}_{\Omega_i}(y_i, z_i) &= \begin{cases} R \times \{0\}, & y_i = 0, z_i > 0, \\ \{0\} \times R, & y_i > 0, z_i = 0, \\ (R_- \times R_-), & y_i = 0, z_i = 0. \end{cases} \end{aligned}$$

进而有

$$\begin{aligned} \hat{N}_{R^n \times \Omega}(\bar{x}_i, \bar{y}_i, \bar{z}_i) &= \{0\} \times \bigotimes_{i=1}^m \hat{N}_{\Omega_i}(\bar{y}_i, \bar{z}_i) = \{(0, d_i^a, d_i^b) : d_\alpha^b = 0, d_\gamma^a = 0, d_\beta^a \leq 0, d_\beta^b \leq 0\}. \\ N_{R^n \times \Omega}(\bar{x}_i, \bar{y}_i, \bar{z}_i) &= \{0\} \times \bigotimes_{i=1}^m N_{\Omega_i}(\bar{y}_i, \bar{z}_i) \\ &= \{(0, d_i^a, d_i^b) : d_\alpha^b = 0, d_\gamma^a = 0, d_\beta^a \leq 0, d_\beta^b \leq 0 \text{ or } d_\beta^a = 0 \text{ or } d_\beta^b = 0\}. \end{aligned}$$

在基本约束规范成立的条件下，抽象约束集合的切锥、法锥形式具体如下：

定理 3.2 [8] MPCC (3.6)的可行域为 $\Phi = \{(x, y, z) \in R^n \times \Omega : F(x, y, z) = 0\}$ ，其中 $R^n \times \Omega \subset R^{n+2m}$ 是一个闭集合， $F: R^{2n} \rightarrow R^p$ 是一个光滑连续可微映射， $F = (g_1, \dots, g_m, h_1, \dots, h_m)^T$ ， $p = 2m$ ，则：

$$1) \quad \hat{N}_\Phi(\bar{x}, \bar{y}, \bar{z}) \supset \left\{ \sum_{i=1}^m \lambda_i \nabla g_i(\bar{x}) + \sum_{i=1}^m \lambda_i \nabla h_i(\bar{x}) + d \mid (\lambda^g, \lambda^h) \in R^{2m}, d \in \hat{N}_D(\bar{x}, \bar{y}, \bar{z}) \right\} \\ = JF(\bar{x}, \bar{y}, \bar{z})^T R^p + \hat{N}_D(\bar{x}, \bar{y}, \bar{z})$$

2) 如果基本约束规范在 $(\bar{x}, \bar{y}, \bar{z})$ 处成立，即：

$$\begin{cases} y \in R^p \\ 0 \in JF(\bar{x}, \bar{y}, \bar{z})^T y + N_{R^n \times \Omega}(\bar{x}, \bar{y}, \bar{z}) \end{cases} \rightarrow y = (0, \dots, 0)^T$$

则 $N_\Phi(\bar{x}, \bar{y}, \bar{z}) \subset JF(\bar{x}, \bar{y}, \bar{z})^T R^p + N_D(\bar{x}, \bar{y}, \bar{z})$ 。

3) 当 $\beta = \emptyset$ 时， $D = R^n \times \Omega$ 在 $(\bar{x}, \bar{y}, \bar{z})$ 处正则， $\{0\}^p$ 在 $F(\bar{x}, \bar{y}, \bar{z})$ 处正则，则 Φ 在 $(\bar{x}, \bar{y}, \bar{z})$ 处正则，得 $\hat{N}_\Phi(\bar{x}, \bar{y}, \bar{z}) = N_\Phi(\bar{x}, \bar{y}, \bar{z}) = JF(\bar{x}, \bar{y}, \bar{z})^T R^p + N_D(\bar{x}, \bar{y}, \bar{z})$ 。

下面给出在基本约束规范成立的条件下，MPCC 问题的 W-稳定点概念。

定理 3.3 设 $\bar{x}$ 为 MPCC (3.6)的局部最优解，且该点处基本约束规范成立，当 $\beta = \emptyset$ 时，严格互补松弛条件成立，则 $\bar{x}$ 为 W-稳定点。即存在 $(\lambda^H, \lambda^G) \in R^{q+p}$ ，满足

$$\nabla f(\bar{x}) + \sum_{i \in \alpha \cup \beta} \lambda_i^H H(\bar{x}) + \sum_{i \in \beta \cup \gamma} \lambda_i^G G(\bar{x}) = 0, \\ \lambda_\gamma^H = 0, \lambda_\alpha^G = 0.$$

证明 由定理 3.2 可知，MPCC 问题的可行域集合的法锥形式如下：

$$N_\Phi(\bar{x}, \bar{y}, \bar{z}) \subseteq JF(\bar{x}, \bar{y}, \bar{z})^T R^p + N_D(\bar{x}, \bar{y}, \bar{z}) \\ \subseteq \begin{pmatrix} H(\bar{x})^T \lambda^H + G(\bar{x})^T \lambda^G \\ -\lambda^H \\ -\lambda^G \end{pmatrix} + \begin{pmatrix} 0 \\ d^H \\ d^G \end{pmatrix} \\ = \begin{pmatrix} H(\bar{x})^T \lambda^H + G(\bar{x})^T \lambda^G \\ -\lambda^H + d^H \\ -\lambda^G + d^G \end{pmatrix}.$$

因为 $\hat{N}_\Phi(\bar{x}, \bar{y}, \bar{z}) \subset N_\Phi(\bar{x}, \bar{y}, \bar{z})$ ，由基本最优性条件可得

$$0 \in \nabla f(\bar{x}) + \hat{N}_\Phi(\bar{x}, \bar{y}, \bar{z}) \\ \in \nabla f(\bar{x}) + N_\Phi(\bar{x}, \bar{y}, \bar{z}) \\ \in \begin{pmatrix} \nabla f(\bar{x}) \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} H(\bar{x})^T \lambda^H + G(\bar{x})^T \lambda^G \\ -\lambda^H + d^H \\ -\lambda^G + d^G \end{pmatrix} \\ = \begin{pmatrix} \nabla f(\bar{x}) + H(\bar{x})^T \lambda^H + G(\bar{x})^T \lambda^G \\ -\lambda^H + d^H \\ -\lambda^G + d^G \end{pmatrix}.$$

所以存在 λ^H, λ^G , 有

$$0 = \nabla f(\bar{x}) + H(\bar{x})^T \lambda^H + G(\bar{x})^T \lambda^G,$$

$$0 = -\lambda^H + d^H,$$

$$0 = -\lambda^G + d^G.$$

由定理 3.1 可知, 当 $i \in \gamma$ 时, 我们有 $d_\gamma^H = 0$, 那么如果 $0 = -\lambda_\gamma^H + d_\gamma^H$, 有 $\lambda_\gamma^H = 0$; 当 $i \in \alpha$ 时, 我们有 $d_\alpha^G = 0$, 那么如果 $0 = -\lambda_\alpha^G + d_\alpha^G$, 有 $\lambda_\alpha^G = 0$.

由此可知 $\bar{x}$ 是 W-稳定点。

4. 增广拉格朗日方法

对于等式约束优化问题

$$\begin{aligned} \min f(x) \\ \text{s.t. } g(x) = 0, x \in R^n. \end{aligned}$$

对这一等式约束优化问题, Powell [9]、Hestenes [10] 在 1969 年提出用增广拉格朗日方法来求解, 首先将上述问题转化为

$$\min f(x) + \lambda^T g(x) + c \|g(x)\|^2$$

的无约束优化问题, 这里 $c > 0$ 是罚参数, λ 是增广拉格朗日乘子的近似。该方法可以通过对 λ 进行迭代, 得到近似最优解。

本节首先将问题(3.6)中的互补约束表示为拉格朗日函数的形式, 它的拉格朗日函数 $L: R^n \times R^p \times R^p \rightarrow R$ 为

$$L(x, y, z, \lambda) = f(x) + \lambda^T F(x, y, z), \quad (4.1)$$

其中关于 x, y, z 的梯度和 Hessian 矩阵如下:

$$\nabla_x L(x, y, z, \lambda) = \nabla f(x) + J_x F(x, y, z)^T \lambda,$$

$$\nabla_y L(x, y, z, \lambda) = J_y F(x, y, z)^T \lambda,$$

$$\nabla_z L(x, y, z, \lambda) = J_z F(x, y, z)^T \lambda.$$

问题(3.6)的增广拉格朗日函数 $L_c: R^n \times R^p \times R^p \rightarrow R$ 为

$$L_c(x, y, z, \lambda) = f(x) + \lambda^T F(x, y, z) + \frac{c}{2} \|F(x, y, z)\|^2, \quad (4.2)$$

其中 $c > 0$ 是一个罚参数, $\|\cdot\|$ 表示欧几里得 l_2 -范数。则上述增广拉格朗日对偶问题为

$$\max_{\lambda} \left\{ \inf_{(x, y, z)} L_c(x, y, z, \lambda^k) \right\}. \quad (4.3)$$

求解该对偶问题的增广拉格朗日方法迭代格式如下:

$$(x^{k+1}, y^{k+1}, z^{k+1}) \in \arg \max_{(x, y, z) \in \Omega} L_c(x, y, z, \lambda^k),$$

$$\lambda^{k+1} = \lambda^k + c^k F(x^{k+1}, y^{k+1}, z^{k+1}).$$

定理 4.1 [11] 设 $(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda})$ 为 $L(x, y, z, \lambda)$ 的一个稳定点, 则对任意的 $c > 0$, $(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda})$ 也是

$L_c(x, y, z, \lambda)$ 的一个稳定点, 而且 $L_c(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda}) = L(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda})$ 。反之也成立。

证明 $L_c(x, y, z, \lambda)$ 的梯度表达式为

$$\nabla_x L_c(x, y, z, \lambda) = \nabla f(x) + J_x F(x, y, z)^T \lambda + c J_x F(x, y, z)^T F(x, y, z), \quad (4.4)$$

$$\nabla_y L_c(x, y, z, \lambda) = J_y F(x, y, z)^T \lambda + c J_y F(x, y, z)^T F(x, y, z), \quad (4.5)$$

$$\nabla_z L_c(x, y, z, \lambda) = J_z F(x, y, z)^T \lambda + c J_z F(x, y, z)^T F(x, y, z), \quad (4.6)$$

$$\nabla_\lambda L_c(x, y, z, \lambda) = F(x, y, z). \quad (4.7)$$

如果 $(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda})$ 是 $L(x, y, z, \lambda)$ 的一个稳定点, 那么有

$$\nabla_x L(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda}) = 0, \nabla_y L(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda}) = 0, \nabla_z L(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda}) = 0, F(\bar{x}, \bar{y}, \bar{z}) = 0.$$

因此, 由(4.4) (4.5) (4.6) (4.7)可知, 对任意的 $c > 0$, 有

$$\nabla_x L_c(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda}) = \nabla f(x) + J_x F(x, y, z)^T \lambda + c J_x F(x, y, z)^T F(x, y, z) = 0,$$

$$\nabla_y L_c(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda}) = J_y F(x, y, z)^T \lambda + c J_y F(x, y, z)^T F(x, y, z) = 0,$$

$$\nabla_z L_c(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda}) = J_z F(x, y, z)^T \lambda + c J_z F(x, y, z)^T F(x, y, z) = 0,$$

$$\nabla_\lambda L_c(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda}) = F(x, y, z) = 0.$$

即 $(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda})$ 是 $L_c(x, y, z, \lambda)$ 的一个稳定点。进一步, 把 $F(\bar{x}, \bar{y}, \bar{z}) = 0$ 代入(4.2)即得

$$L_c(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda}) = L(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda}).$$

同样的, 当

$$\nabla_x L_c(\bar{x}, \bar{y}, \bar{z}, \lambda) = 0, \nabla_y L_c(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda}) = 0, \nabla_z L_c(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda}) = 0, \nabla_\lambda L_c(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda}) = 0,$$

有

$$\nabla_x L(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda}) = 0, \nabla_y L(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda}) = 0, \nabla_z L(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda}) = 0, F(\bar{x}, \bar{y}, \bar{z}) = 0.$$

这说明 $(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda})$ 是 $L(x, y, z, \lambda)$ 的一个稳定点。把 $F(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda}) = 0$ 代入(4.2)就可以得到

$$L_c(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda}) = L(\bar{x}, \bar{y}, \bar{z}, \bar{\lambda}).$$

最后, 我们证明采用增广拉格朗日方法用于求解该等价 MPCC 问题所生成的点列收敛到该等价问题的 W-稳定点。

定理 4.2 设 $\{(x^k, y^k, z^k)\}$ 为增广拉格朗日对偶问题(4.3)的 KKT 稳定点序列, 假设 $(\bar{x}, \bar{y}, \bar{z}) \in \Phi$ 是它的聚点, 在 $\bar{x}$ 处 MPCC-LICQ 成立, 则当 $\{c_k\} \rightarrow 0$ 时, $\bar{x}$ 是问题 MPCC (3.6)的 W-稳定点。

证明 因为 $\{(x^k, y^k, z^k)\}$ 为问题(4.3)的 KKT 稳定点序列, 那么

$$\nabla_x L_{c^k}(x^k, y^k, z^k, \lambda^k) = \nabla f(x^k) + J_{x^k} F(x^k, y^k, z^k)^T \lambda^k + c^k J_{x^k} F(x^k, y^k, z^k)^T F(x^k, y^k, z^k) = 0.$$

记 $\lambda^k = (\lambda_H^k, \lambda_G^k)$,

$$\nabla_x L_{c^k}(x^k, y^k, z^k, \lambda^k) = \nabla f(x^k) + \begin{pmatrix} H(x^k) \\ G(x^k) \end{pmatrix}^T \lambda^k + c^k \begin{pmatrix} H(x^k)^T & G(x^k)^T \end{pmatrix} \begin{pmatrix} A^T(Ax - b) + Jg(x)^T y \\ g(x) + y - z \end{pmatrix} = 0.$$

可以推出

$$\nabla f(x^k) + H(x^k)^T \lambda_H^k + G(x^k)^T \lambda_G^k + c^k \left(H(x^k)^T (A^T (Ax - b) + Jg(x)^T y) + G(x^k)^T (g(x) + y - z) \right) = 0.$$

由上式可得

$$\begin{aligned} -\nabla f(x^k) &= H(x^k)^T \lambda_H^k + G(x^k)^T \lambda_G^k + c^k \left(H(x^k)^T (A^T (Ax - b) + Jg(x)^T y) + G(x^k)^T (g(x) + y - z) \right) \\ &= H(x^k)^T \left(\lambda_H^k + c^k (A^T (Ax - b) + Jg(x)^T y) \right) + G(x^k)^T (\lambda_G^k + c^k (g(x) + y - z)). \end{aligned}$$

令

$$a^k = \lambda_H^k + c^k (A^T (Ax - b) + Jg(x)^T y), b^k = \lambda_G^k + c^k (g(x) + y - z), \quad (4.8)$$

即

$$-\nabla f(x^k) = H(x^k)^T a^k + G(x^k)^T b^k = \sum_{i=1}^p a_i^k H(x^k) + \sum_{i=1}^p b_i^k G(x^k). \quad (4.9)$$

下证式(4.8)中的乘子序列 $\{a^k, b^k\}$ 有界。如果 $\{a^k, b^k\}$ 无界, 那么存在一个子集 K , 对任意的 $k \in K$,

$k \rightarrow \infty$, 有 $\frac{(a^k, b^k)}{\|(a^k, b^k)\|} \rightarrow (a', b')$ 。(4.9)式两边同时除以 $\|(a^k, b^k)\|$ 并取极限, 当 $k \rightarrow \infty$ 时, 有

$$\sum_{i \in \alpha \cup \beta} a_i' H(\bar{x}) + \sum_{i \in \beta \cup \gamma} b_i' G(\bar{x}) = 0.$$

当 $i \in \gamma$ 时,

$$\lim_{k \rightarrow \infty} a^k = \lambda_H^k + c^k (A^T (Ax - b) + Jg(x)^T y) = 0.$$

当 $i \in \alpha$ 时,

$$\lim_{k \rightarrow \infty} b^k = \lambda_G^k + c^k (g(x) + y - z) = 0.$$

因为 $(a', b') \neq 0$ 与在 $\bar{x}$ 处满足 MPCC-LICQ 矛盾, 所以 $\{(a^k, b^k)\}$ 有界。

不失一般性, 设 $\{(a^k, b^k)\}$ 收敛于 $(\bar{a}, \bar{b})$ 设

$$\lim_{k \rightarrow \infty} a_i^k = \bar{a}_i, \lim_{k \rightarrow \infty} b_i^k = \bar{b}_i, i = 1, \dots, m.$$

当 $i \in \gamma$ 时, $\lambda_H^k = 0$, $k \rightarrow \infty$, $c^k \rightarrow 0$ 时, 即 $\bar{a}_i = 0$ 。同理, 当 $i \in \alpha$ 时, 可得 $\bar{b}_i = 0$ 。由 H, G 连续可微, 对(4.9)取极限, 得

$$\begin{aligned} -\nabla f(\bar{x}) &= \sum_{i \in \alpha \cup \beta} \bar{a}_i H(\bar{x}) + \sum_{i \in \beta \cup \gamma} \bar{b}_i G(\bar{x}), \\ \nabla f(\bar{x}) &= 0. \end{aligned}$$

即证得 $\bar{x}$ 是问题 MPCC(3.6)的 W-稳定点。

5. 结论

本文主要针对约束不相容的凸优化问题建立了最小约束违背优化模型。当约束相容时, 模型退化为原始问题。当约束不相容时, 模型可被重新表述为 MPCC 问题。并证明了该等价 MPCC 问题的 W-稳定性。将增广拉格朗日方法用于求解该等价问题, 证明了该方法生成的点列收敛到该等价问题的 W-稳定点。

对于最小约束违背优化, 还有许多问题值得研究。当不知道可行域是否为空集时, 最小约束违背优

化问题总是可行的，这是它的优点。如果约束问题是可行的，那么问题涉及的不可行性度量 $\theta(x)$ 通常是光滑的，但不是二次可微的，这会带来计算上的困难。本文只处理了最小约束违背的凸优化问题，增广拉格朗日方法能否处理最小约束违背的非凸优化问题需要继续思考和研究。

参考文献

- [1] Byrd, R.H., Curtis, F.E. and Nocedal, J. (2010) Infeasibility Detection and SQP Methods for Nonlinear Optimization. *SIAM Journal on Optimization*, **20**, 2281-2299. <https://doi.org/10.1137/080738222>
- [2] Burke, J.V., Curtis, F.E. and Wang, H. (2014) A Sequential Quadratic Optimization Algorithm with Rapid Infeasibility Detection. *SIAM Journal on Optimization*, **24**, 839-872. <https://doi.org/10.1137/120880045>
- [3] Dai, Y., Liu, X. and Sun, J. (2020) A Primal-Dual Interior-Point Method Capable of Rapidly Detecting Infeasibility for Nonlinear Programs. *Journal of Industrial & Management Optimization*, **16**, 1009-1035. <https://doi.org/10.3934/jimo.2018190>
- [4] 戴戡虹, 张立卫. 最小约束违背优化[M]. 北京: 科学出版社, 2023: 223-253.
- [5] Dai, Y.-H. and Zhang, L. (2021) Optimization with Least Constraint Violation. *CSIAM Transactions on Applied Mathematics*, **2**, 551-584. <https://doi.org/10.4208/csiam-am.2020-0043>
- [6] Chiche, A. and Gilbert, J.C.H. (2016) How the Augmented Lagrangian Algorithm Can Deal with An Infeasible Convex Quadratic Optimization Problem. *Journal of Convex Analysis*, **23**, 425-459.
- [7] Rockafellar, R.T. and Wets, R.J.-B. (1998) Variational Analysis. Springer-Verlag.
- [8] 张立卫, 吴佳, 张艺. 变分分析[M]. 北京: 科学出版社, 2013: 100-101.
- [9] Powell, M.J.D. (1969) A Method for Nonlinear Constraints in Minimization Problems. In: Fletcher, R., Ed., *Optimization*, Academic Press, 283-298.
- [10] Hestenes, M.R. (1969) Multiplier and Gradient Methods. *Journal of Optimization Theory and Applications*, **4**, 303-320. <https://doi.org/10.1007/bf00927673>
- [11] 刘昶. 互补约束优化问题的增广拉格朗日方法[D]: [硕士学位论文]. 大连: 大连理工大学, 2022.