

基于 BB 步长的一类原始-对偶算法^①

郑代秀

四川师范大学 数学科学学院, 成都 610066

摘要: 增广拉格朗日乘子法(ALM)是求解带等式约束的二次凸优化问题的常用方法, 但罚参数选取不当时, 收敛速度比较慢. 提出 ALM-BB 算法, 利用 Barzilai-Borwein(BB)算法的步长去改进原始的 ALM, 证明 ALM-BB 算法的收敛性. 最后将这类方法运用于求解范数最优控制问题. 数值算例表明改进的算法收敛速度更快.

关键词: 二次凸规划问题; 增广拉格朗日乘子法; BB 步长; 范数最优控制问题

中图分类号: O232

文献标志码: A

文章编号: 1000-5471(2023)01-0040-08

A Class of Primal-Dual Algorithms Based on BB Step Size

ZHENG Daixiu

College of Mathematical Sciences, Sichuan Normal University, Chengdu 610066, China

Abstract: A class of numerical computational methods for quadratic convex optimization problems with equational constraints is studied. The Augmented Lagrange Multiplier Method (ALM) is a common method for solving such problems, but the convergence speed is slow due to the improper selection of penalty parameters. We propose to improve the original ALM algorithm by using the step size of Barzilai-Borwein (BB) algorithm to prove the convergence of ALM-BB algorithm. Finally, this method is applied to solve the parametric optimal control problem. Numerical examples show that the improved algorithm converges faster.

Key words: quadratic convex programming problem; generalized Lagrange multiplier method; BB step size; parametric optimal control problem

对于带线性等式约束的凸优化问题

$$\min\{f(\mathbf{x}) \mid \mathbf{Ax} - \mathbf{c} = \mathbf{0}\} \quad (1)$$

其中 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 为强凸二次函数, $\mathbf{c} \in \mathbb{R}^m$, $\mathbf{A} \in \mathbb{R}^{m \times n}$ 且 $\text{rank}(\mathbf{A}) = m$ ($n, m \in \mathbb{N}$, $m < n$), 其中 $\text{rank}(\mathbf{A})$ 表示矩阵 $\mathbf{A}$ 的秩. 求解带等式约束的优化问题的一个经典方法是外罚函数法: 通过增加惩罚项将约束问题转为带罚参数的无约束优化问题, 并通过求解无约束优化问题得到原约束优化问题的近似解. 外罚函数法

① 收稿日期: 2021-12-31

基金项目: 国家自然科学基金面上项目(12071324).

作者简介: 郑代秀, 硕士研究生, 主要从事运筹学与控制论研究.

最早是由文献[1]提出. 在外罚函数法中, 罚参数选取过小时, 罚函数可能无界, 此时迭代发散^[2]; 而当罚参数趋于无穷时, 增广目标函数将呈现病态性质^[3]. 为此, 文献[4]和文献[5]于 1969 年各自独立地提出在罚函数中引入 Lagrange 乘子项来改进算法, 该算法被称为增广 Lagrange 乘子法(ALM). 解实际问题时, 与外罚函数法相比, ALM 中罚因子 ρ 是可以固定的, 而不需要让其趋于无穷^[6]. 但罚参数仍需适当选取, 若罚参数 ρ 选取过大, 增广 Lagrange 函数关于 $\mathbf{x}$ 的海瑟矩阵的条件数会病态, 从而使得求解难度提高. 在文献[2]中, 提到对罚参数 ρ 的一个经验选取是 $\rho \in [2, 10]$. 文献[7]建议, 用一个单调递增的序列 $\{\rho_k\}$ 代替固定的 ρ , 但也没有从根本上解决罚参数选取难的问题.

ALM 算法本质上是一类原始-对偶算法, 更新乘子 λ 的过程本质上是用梯度法求解对偶问题, 其中参数 ρ 是梯度算法的下降步长. 而在优化问题的梯度类算法中, Barzilai-Borwein 算法(以下简称 BB 算法^[8])的表现比较好(具有拟牛顿性质). 所以可以考虑用 BB 算法的步长去替代 ALM 中固定步长 ρ , 得到一个新的凸优化问题的原始-对偶算法, 我们将其简记为 ALM-BB 算法. 该思想首先由文献[9]在研究大数值优化问题的乘子交替方向法^[10-11]的改进型算法时提出, 但没给出算法的收敛性证明. 本文在目标函数 f 为强凸二次函数, 约束函数中矩阵 $\mathbf{A}$ 行满秩的条件下证明了 ALM-BB 算法收敛性.

范数最优控制问题本质上是无穷维的二次凸优化问题, 通过适当的离散格式可以获得其有限维框架的近似问题, 且其近似问题为一类二次凸优化问题. 由此可以利用本文的算法进行数值求解. 通过求解范数最优控制问题的数值算例发现, ALM-BB 算法收敛速度更快.

本文的结构如下: 在第 1 节中给出一些记号和预备知识; 在第 2 节中介绍 ALM-BB 算法并证明其收敛性; 在第 3 节中给出范数最优控制的离散格式, 并利用实际算例说明 ALM-BB 算法的有效性; 在第 4 节中作总结.

1 预备知识

考虑如下二次凸优化问题:

$$\begin{aligned} \min \quad & \frac{1}{2} \mathbf{x}^T \mathbf{S} \mathbf{x} - \mathbf{b}^T \mathbf{x} \\ \text{s. t.} \quad & \mathbf{A} \mathbf{x} - \mathbf{c} = \mathbf{0} \end{aligned} \quad (\text{P})$$

其中 $\mathbf{S} \in \mathbb{R}^{n \times n}$, $\mathbf{b} \in \mathbb{R}^n$, $\mathbf{c} \in \mathbb{R}^m$, $\mathbf{A} \in \mathbb{R}^{m \times n}$. 在本文中, 我们作出如下假设:

(A1) $\mathbf{S} \in \mathbb{R}^{n \times n}$ 为对称正定矩阵;

(A2) $\text{rank}(\mathbf{A}) = m$.

问题(P) 的 Lagrange 函数 $\mathcal{L}: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ 定义为

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \frac{1}{2} \mathbf{x}^T \mathbf{S} \mathbf{x} - \mathbf{b}^T \mathbf{x} - \boldsymbol{\lambda}^T (\mathbf{A} \mathbf{x} - \mathbf{c})$$

向量 $\boldsymbol{\lambda}$ 称为对偶变量或者原问题(P) 的 Lagrange 乘子向量.

Lagrange 对偶函数定义为

$$d(\boldsymbol{\lambda}) = \inf_{\mathbf{x} \in \mathbb{R}^n} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$$

问题(P) 的对偶问题定义为

$$\max_{\boldsymbol{\lambda} \in \mathbb{R}^m} d(\boldsymbol{\lambda}) \quad (\text{D})$$

引理 1^[2] 对于凸优化问题(P), 若条件(A1)–(A2) 成立, 那么 $\mathbf{x}^*$, $\boldsymbol{\lambda}^*$ 分别是原、对偶问题的全局最优解当且仅当

$$\begin{cases} \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0} \\ \nabla_{\boldsymbol{\lambda}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0} \end{cases} \quad (2)$$

称满足(2) 式的 $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ 为 KKT 对. 由引理 1 知求解凸优化问题(P) 的 KKT 对等价于求解原问题与

对偶问题的最优解.

优化问题(P) 的增广 Lagrange 函数为:

$$\mathcal{L}_\rho(\mathbf{x}, \boldsymbol{\lambda}) = \frac{1}{2} \mathbf{x}^\top \mathbf{S} \mathbf{x} - \mathbf{b}^\top \mathbf{x} - \langle \boldsymbol{\lambda}, \mathbf{A} \mathbf{x} - \mathbf{c} \rangle + \frac{\rho}{2} \|\mathbf{A} \mathbf{x} - \mathbf{c}\|^2 \quad (3)$$

其中 $\rho > 0$ 是罚参数. 同样, 可利用其增广 Lagrange 函数定义对偶函数:

$$d_\rho(\boldsymbol{\lambda}) = \inf_{\mathbf{x} \in \mathbb{R}^n} \mathcal{L}_\rho(\mathbf{x}, \boldsymbol{\lambda})$$

引理 2^[12] 若 $\mathbf{x}^*, \boldsymbol{\lambda}^*$ 分别为增广 Lagrange 函数 $\mathcal{L}_\rho(\mathbf{x}, \boldsymbol{\lambda})$ 对应的原始、对偶问题的解, 则 $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ 为 Lagrange 函数 $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$ 的 KKT 对.

下面我们回顾经典的增广 Lagrange 乘子算法. 求解问题(P) 的增广 Lagrange 乘子法(ALM) 的迭代格式如下:

算法 1: ALM

1. 给定初始点 $\mathbf{x}_0$ 与 $\boldsymbol{\lambda}_0$, $\rho > 0$, 终止误差 ϵ , 令 $k = 0$;
 2. 计算 $\mathbf{x}_{k+1} = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \mathcal{L}_\rho(\mathbf{x}, \boldsymbol{\lambda}_k)$;
 3. 若 $\|\mathbf{A} \mathbf{x}_{k+1} - \mathbf{c}\| < \epsilon$, 停止; 否则转第 4 步;
 4. 计算 $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k - \rho(\mathbf{A} \mathbf{x}_{k+1} - \mathbf{c})$;
 5. 令 $k = k + 1$, 转第 2 步.
-

该算法中的步长 ρ 是固定的, 而罚参数若选取不当, 会显著增加求解时间. 为改进算法, 考虑用 BB 算法的步长 α_{BB} 去替代固定的步长 ρ . 给定 $\mathbf{x}_k, \mathbf{x}_{k-1}, \boldsymbol{\lambda}_k, \boldsymbol{\lambda}_{k-1}$, 定义

$$\alpha_{\text{BB}}^k = \frac{\langle \mathbf{s}_{k-1}, \mathbf{s}_{k-1} \rangle}{\langle \mathbf{s}_{k-1}, \mathbf{y}_{k-1} \rangle}$$

其中 $\mathbf{s}_{k-1} = \boldsymbol{\lambda}_k - \boldsymbol{\lambda}_{k-1}$, $\mathbf{y}_{k-1} = \mathbf{A} \mathbf{x}_k - \mathbf{A} \mathbf{x}_{k-1}$. 利用 α_{BB}^k 的定义方式, 我们得到如下 ALM-BB 算法.

算法 2: ALM-BB

1. 给定初始点 $\mathbf{x}_0$ 与 $\boldsymbol{\lambda}_0$, $\rho > 0$, 终止误差 ϵ , 令 $k = 0$;
 2. 计算 $\mathbf{x}_{k+1} = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \mathcal{L}_\rho(\mathbf{x}, \boldsymbol{\lambda}_k)$;
 3. 若 $\|\mathbf{A} \mathbf{x}_{k+1} - \mathbf{c}\| < \epsilon$, 停止; 否则转第 4 步;
 4. 若 $k = 0$, 计算 $\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_0 - \rho(\mathbf{A} \mathbf{x}_1 - \mathbf{c})$;
否则计算 $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k - \alpha_{\text{BB}}^k(\mathbf{A} \mathbf{x}_{k+1} - \mathbf{c})$;
 5. 令 $k = k + 1$, 转第 2 步.
-

2 ALM-BB 算法的收敛性

在本节, 我们证明算法 2 的收敛性. 首先, 我们证明算法 2 中更新 Lagrange 乘子的过程等价于求增广 Lagrange 函数对应的对偶问题的 BB 算法.

定理 1 若条件(A1)–(A2) 成立, 则算法 2 中更新 Lagrange 乘子的过程等价于求增广 Lagrange 函数对应的对偶问题的 BB 算法.

证 由条件(A1), 对任意给定的 $\boldsymbol{\lambda} \in \mathbb{R}^m$, 存在唯一的 $\mathbf{x} \in \mathbb{R}^n$ 使得

$$d_\rho(\boldsymbol{\lambda}) = \mathcal{L}_\rho(\mathbf{x}, \boldsymbol{\lambda})$$

由于 $\mathbf{x} \mapsto \mathcal{L}_\rho(\mathbf{x}, \boldsymbol{\lambda})$ 为可微凸函数, 由一阶最优性条件可得:

$$\nabla_{\mathbf{x}} \mathcal{L}_\rho(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{S} \mathbf{x} - \mathbf{b} - \mathbf{A}^\top \boldsymbol{\lambda} + \rho \mathbf{A}^\top (\mathbf{A} \mathbf{x} - \mathbf{c}) = \mathbf{0}$$

由此可得 $\mathbf{x}$ 的显式表达式:

$$\mathbf{x} = \mathbf{R}^{-1} \mathbf{A}^\top \boldsymbol{\lambda} + \mathbf{R}^{-1} \mathbf{h}$$

其中 $\mathbf{R} = \mathbf{S} + \rho \mathbf{A}^\top \mathbf{A}$, $\mathbf{h} = \rho \mathbf{A}^\top \mathbf{c} + \mathbf{b}$. 因此对偶函数

$$\begin{aligned}
d_\rho(\lambda) &= \frac{1}{2}(\mathbf{R}^{-1}\mathbf{A}^\top\lambda + \mathbf{R}^{-1}\mathbf{h})^\top \mathbf{S}(\mathbf{R}^{-1}\mathbf{A}^\top\lambda + \mathbf{R}^{-1}\mathbf{h}) - \mathbf{b}^\top(\mathbf{R}^{-1}\mathbf{A}^\top\lambda + \mathbf{R}^{-1}\mathbf{h}) - \\
&\quad \langle \lambda, \mathbf{A}(\mathbf{R}^{-1}\mathbf{A}^\top\lambda + \mathbf{R}^{-1}\mathbf{h}) - \mathbf{c} \rangle + \frac{\rho}{2} \|\mathbf{A}(\mathbf{R}^{-1}\mathbf{A}^\top\lambda + \mathbf{R}^{-1}\mathbf{h}) - \mathbf{c}\|^2 = \\
&\quad \frac{1}{2}\lambda^\top \mathbf{A}\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1}\mathbf{A}^\top\lambda + \mathbf{h}^\top \mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1}\mathbf{A}^\top\lambda - \mathbf{b}^\top \mathbf{R}^{-1}\mathbf{A}^\top\lambda - \\
&\quad \lambda^\top \mathbf{A}\mathbf{R}^{-1}\mathbf{A}^\top\lambda - \mathbf{h}^\top \mathbf{R}^{-1}\mathbf{A}^\top\lambda + \mathbf{c}^\top\lambda + \\
&\quad \frac{\rho}{2}\lambda^\top \mathbf{A}\mathbf{R}^{-1}\mathbf{A}^\top \mathbf{A}\mathbf{R}^{-1}\mathbf{A}^\top\lambda + \rho \mathbf{h}^\top \mathbf{R}^{-1}\mathbf{A}^\top \mathbf{A}\mathbf{R}^{-1}\mathbf{A}^\top\lambda - \rho \mathbf{c}^\top \mathbf{A}\mathbf{R}^{-1}\mathbf{A}^\top\lambda + C = \\
&\quad \frac{1}{2}\lambda^\top (\mathbf{A}\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1}\mathbf{A}^\top - 2\mathbf{A}\mathbf{R}^{-1}\mathbf{A}^\top + \rho \mathbf{A}\mathbf{R}^{-1}\mathbf{A}^\top \mathbf{A}\mathbf{R}^{-1}\mathbf{A}^\top)\lambda - \\
&\quad (-\mathbf{h}^\top \mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1}\mathbf{A}^\top + \mathbf{b}^\top \mathbf{R}^{-1}\mathbf{A}^\top + \mathbf{h}^\top \mathbf{R}^{-1}\mathbf{A}^\top - \mathbf{c}^\top - \rho \mathbf{h}^\top \mathbf{R}^{-1}\mathbf{A}^\top \mathbf{A}\mathbf{R}^{-1}\mathbf{A}^\top + \\
&\quad \rho \mathbf{c}^\top \mathbf{A}\mathbf{R}^{-1}\mathbf{A}^\top)\lambda + C = \\
&\quad -\frac{1}{2}\lambda^\top \mathbf{A}\mathbf{R}^{-1}\mathbf{A}^\top\lambda - (\mathbf{h}^\top \mathbf{R}^{-1}\mathbf{A}^\top - \mathbf{c}^\top)\lambda + C
\end{aligned} \tag{4}$$

其中 C 为与 $\mathbf{x}, \lambda$ 无关的常数项.

由(4)式, 增广 Lagrange 函数对应的对偶问题可写为如下无约束凸二次优化问题(为简单起见, 我们将常数项 C 省略, 它对对偶问题最优解的求解不会产生影响):

$$\min_{\lambda \in \mathbb{R}^m} -\tilde{d}_\rho(\lambda) = \frac{1}{2}\lambda^\top \mathbf{A}\mathbf{R}^{-1}\mathbf{A}^\top\lambda + (\mathbf{h}^\top \mathbf{R}^{-1}\mathbf{A}^\top - \mathbf{c}^\top)\lambda \tag{5}$$

在假设条件(A1)–(A2)下, 矩阵 $\mathbf{A}\mathbf{R}^{-1}\mathbf{A}^\top$ 是对称正定矩阵, 因此优化问题(5)的解存在唯一.

由文献[8], 求解无约束凸优化问题(5)的 BB 算法迭代格式为

$$\lambda_{k+1} = \lambda_k - \tilde{\alpha}_{\text{BB}}^k \mathbf{g}(\lambda_k) \tag{6}$$

这里 $\tilde{\alpha}_{\text{BB}}^k = \frac{\langle \mathbf{s}_{k-1}, \mathbf{s}_{k-1} \rangle}{\langle \mathbf{s}_{k-1}, \mathbf{y}_{\lambda_{k-1}} \rangle}$, $\mathbf{s}_{k-1} = \lambda_k - \lambda_{k-1}$, $\mathbf{y}_{\lambda_{k-1}} = \mathbf{g}(\lambda_k) - \mathbf{g}(\lambda_{k-1})$, $\mathbf{g}(\lambda_k) = \nabla(-\tilde{d}_\rho(\lambda_k)) = \mathbf{A}\mathbf{R}^{-1}\mathbf{A}^\top\lambda_k + \mathbf{A}\mathbf{R}^{-1}\mathbf{h} - \mathbf{c}$.

在 ALM-BB 算法第 2 步中, 利用凸优化问题的一阶最优性条件可得 $\mathbf{x}_{k+1}$ 的精确解为

$$\mathbf{x}_{k+1} = \mathbf{R}^{-1}\mathbf{A}^\top\lambda_k + \mathbf{R}^{-1}\mathbf{h} \tag{7}$$

将(7)式代入 ALM-BB 算法第 3 步, 则有

$$\lambda_{k+1} = \lambda_k - \alpha_{\text{BB}}^k (\mathbf{A}(\mathbf{R}^{-1}\mathbf{A}^\top\lambda_k + \mathbf{R}^{-1}\mathbf{h}) - \mathbf{c})$$

对比算法 2 与对偶问题的 BB 算法, 我们只需证明

$$\mathbf{g}(\lambda_k) = \nabla(-\tilde{d}_\rho(\lambda_k)) = \mathbf{A}\mathbf{x}_{k+1} - \mathbf{c}$$

其中 $\mathbf{x}_{k+1}$ 为算法 2 中求得的子问题的最优解.

事实上, 由凸优化问题的一阶最优性条件可知 $\mathbf{x}_{k+1}$ 应满足

$$\nabla_{\mathbf{x}} \mathcal{L}_\rho(\mathbf{x}_{k+1}, \lambda_k) = \mathbf{S}\mathbf{x}_{k+1} - \mathbf{b} - \mathbf{A}^\top\lambda_k + \rho \mathbf{A}^\top(\mathbf{A}\mathbf{x}_{k+1} - \mathbf{c}) = \mathbf{0}$$

利用之前的记号可得

$$\mathbf{x}_{k+1} = \mathbf{R}^{-1}\mathbf{A}^\top\lambda_k + \mathbf{R}^{-1}\mathbf{h} \tag{8}$$

故由(5)式中 $-\tilde{d}_\rho(\lambda)$ 的定义, 有

$$\begin{aligned}
\nabla(-\tilde{d}_\rho(\lambda)) &= \mathbf{A}\mathbf{R}^{-1}\mathbf{A}^\top\lambda_k + \mathbf{A}\mathbf{R}^{-1}\mathbf{h} - \mathbf{c} = \\
&\quad \mathbf{A}(\mathbf{R}^{-1}\mathbf{A}^\top\lambda_k + \mathbf{R}^{-1}\mathbf{h}) - \mathbf{c} = \\
&\quad \mathbf{A}\mathbf{x}_{k+1} - \mathbf{c}
\end{aligned}$$

下面我们利用 BB 算法的收敛性结果证明算法 2 产生的迭代序列的收敛性.

定理 2 设条件(A1) – (A2) 成立, 则由算法 2 得到的迭代序列 $\{(\mathbf{x}_k, \boldsymbol{\lambda}_k)\}$ 收敛到原问题的 KKT 对 $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$. 特别地, $\mathbf{x}^*$ 为原问题的最优解.

证 由条件(A1) – (A2) 知 $\mathbf{x}^*$ 为原问题的最优解等价于存在 $\boldsymbol{\lambda}^*$ 使得 $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ 为 KKT 对, 即 $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ 满足

$$\begin{cases} \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \mathbf{S}\mathbf{x}^* - \mathbf{b} - \mathbf{A}^T \boldsymbol{\lambda}^* = \mathbf{0} \\ \nabla_{\boldsymbol{\lambda}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \mathbf{A}\mathbf{x}^* - \mathbf{c} = \mathbf{0} \end{cases}$$

根据 BB 算法的收敛性(见文献[13] 定理 2.5) 及定理 1, 有 $\boldsymbol{\lambda}_k \rightarrow \boldsymbol{\lambda}^*, k \rightarrow \infty$. 其中 $\boldsymbol{\lambda}^*$ 为对偶问题(5) 的最优解. 因此, 由一阶最优性条件,

$$\nabla(-\tilde{d}_\rho(\boldsymbol{\lambda}^*)) = \mathbf{A}(\mathbf{R}^{-1}\mathbf{A}^T\boldsymbol{\lambda}^* + \mathbf{R}^{-1}\mathbf{h}) - \mathbf{c} = \mathbf{0} \quad (9)$$

另一方面, 由 $\mathbf{x}_{k+1}$ 的精确表达式(8) 及 $\{\boldsymbol{\lambda}_k\}$ 的收敛性知 $\mathbf{x}_{k+1} \rightarrow \mathbf{x}^*, k \rightarrow \infty$, 其中

$$\mathbf{x}^* = \mathbf{R}^{-1}\mathbf{A}^T\boldsymbol{\lambda}^* + \mathbf{R}^{-1}\mathbf{h} \quad (10)$$

结合(9) 式和(10) 式可得

$$\mathbf{A}\mathbf{x}^* = \mathbf{A}\mathbf{R}^{-1}\mathbf{A}^T\boldsymbol{\lambda}^* + \mathbf{A}\mathbf{R}^{-1}\mathbf{h} = \mathbf{c}$$

即 $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ 满足条件 $\nabla_{\boldsymbol{\lambda}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}$.

进一步,

$$\begin{aligned} \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*) &= \mathbf{S}\mathbf{x}^* - \mathbf{b} - \mathbf{A}^T \boldsymbol{\lambda}^* = \\ &= \mathbf{S}\mathbf{x}^* - \mathbf{b} - \mathbf{A}^T \boldsymbol{\lambda}^* + \rho \mathbf{A}^T (\mathbf{A}\mathbf{x}^* - \mathbf{c}) = \\ &= (\mathbf{S} + \rho \mathbf{A}^T \mathbf{A})\mathbf{x}^* - \mathbf{A}^T \boldsymbol{\lambda}^* - \mathbf{b} - \rho \mathbf{A}^T \mathbf{c} = \\ &= \mathbf{R}\mathbf{x}^* - \mathbf{A}^T \boldsymbol{\lambda}^* - \mathbf{h} \end{aligned} \quad (11)$$

由(10) 式和(11) 式可得

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \mathbf{R}(\mathbf{R}^{-1}\mathbf{A}^T\boldsymbol{\lambda}^* + \mathbf{R}^{-1}\mathbf{h}) - \mathbf{A}^T \boldsymbol{\lambda}^* - \mathbf{h} = \mathbf{0}$$

故 $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ 为原问题的 KKT 对.

此外, 利用 BB 算法的 R-线性收敛率, 知 ALM-BB 也具有 R-线性收敛率.

3 范数最优控制问题及其离散格式

设 $T > 0$, $\mathbb{R}^n$ 与 $\mathbb{R}^m$ 分别为 n 维与 m 维欧式空间, 考虑如下线性控制系统:

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) + \mathbf{b}(t), \text{ a. e. } t \in [0, T] \\ \mathbf{x}(0) = \mathbf{x}_0, \mathbf{x}(T) = \mathbf{x}_T \end{cases} \quad (12)$$

其中: $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times m}$; $\mathbf{x}(\cdot)$ 为系统的状态, $\mathbf{u}(\cdot)$ 为系统的控制, $\mathbf{x}_0 \in \mathbb{R}^n$ 是给定的初值, $\mathbf{x}_T \in \mathbb{R}^n$ 为给定的终值. 称 $\mathbf{x}(T) = \mathbf{x}_T$ 为终端状态约束. 定义性能指标:

$$J(\mathbf{u}(\cdot)) = \frac{1}{2} \int_0^T \|\mathbf{u}(t)\|^2 dt \quad (13)$$

定义可行控制集 $\mathcal{U} := \{\mathbf{u}(\cdot): [0, T] \longrightarrow \mathbb{R}^m \mid \mathbf{u}(\cdot) \text{ Lebesgue 可测}, \|\mathbf{u}\|_{L^2}^2 = \int_0^T \|\mathbf{u}(t)\|^2 dt < \infty\}$.

带约束的范数最优控制问题是: 寻找控制 $\hat{\mathbf{u}}(\cdot) \in \mathcal{U}$ 及控制系统(12) 相应于 $\hat{\mathbf{u}}(\cdot)$ 的满足约束条件的状态 $\hat{\mathbf{x}}(\cdot)$, 使得

$$J(\hat{\mathbf{u}}(\cdot)) = \inf\{J(\mathbf{u}(\cdot)) \mid \mathbf{u} \in \mathcal{U}, \mathbf{x}(T, \mathbf{x}_0, \mathbf{u}(\cdot)) = \mathbf{x}_T\} \quad (14)$$

若 $\hat{\mathbf{u}}$ 是上述问题的解, 则称 $\hat{\mathbf{u}}$ 为最优控制, 相应的状态 $\hat{\mathbf{x}}$ 称为最优状态.

下面, 我们给出范数最优控制问题(14) 的离散格式. 对控制系统(12) 中的线性微分方程, 我们有 $\mathbf{x}(T)$ 的显示表达式:

$$\mathbf{x}(T) = e^{AT} \mathbf{x}_0 + \int_0^T e^{A(T-t)} \mathbf{B} \mathbf{u}(t) dt$$

记 $\boldsymbol{\eta} = e^{AT} \mathbf{x}_0$, $\mathbf{M}(t) = e^{A(T-t)} \mathbf{B}$. 则状态约束可表示为:

$$\boldsymbol{\eta} + \int_0^T \mathbf{M}(t) \mathbf{u}(t) dt = \mathbf{x}_T \quad (15)$$

将(15) 式中的积分表达式采用分段常量的逼近策略, 取

$$\mathbf{M}(t) = \mathbf{M}(t_i), t \in [t_i, t_{i+1}), i = 0, 1, \dots, N-1$$

其中 $0 = t_0 < t_1 < \dots < t_N = T$ 定义如下:

$$t_0 = 0, t_1 = \Delta t, \dots, t_k = k \Delta t, \dots, t_N = T, \Delta t = \frac{T}{N}$$

可得如下近似方程:

$$\boldsymbol{\eta} + \sum_{i=1}^{N-1} \mathbf{M}(t_i) \mathbf{u}(t_i) \Delta t = \mathbf{x}_T$$

若记 $\mathbf{u} = (\mathbf{u}^T(t_0), \mathbf{u}^T(t_1), \dots, \mathbf{u}^T(t_{N-1}))^T$. 在适当的可积性条件下, 可得:

$$\frac{1}{2} \int_0^T \|\mathbf{u}(t)\|^2 dt \approx \frac{1}{2} \sum_{i=0}^{N-1} \Delta t \mathbf{u}^T(t_i) \mathbf{u}(t_i) = \frac{1}{2} \mathbf{u}^T \mathbf{S} \mathbf{u}$$

其中

$$\mathbf{S} = \begin{pmatrix} \Delta t & & & \\ & \Delta t & & \\ & & \ddots & \\ & & & \Delta t \end{pmatrix}$$

由此可得, 范数最优控制问题(12) 对应的离散优化问题为:

$$\begin{aligned} \min \quad & \frac{1}{2} \mathbf{u}^T \mathbf{S} \mathbf{u} \\ \text{s. t.} \quad & \mathbf{M} \mathbf{u} + \mathbf{c} = \mathbf{0} \end{aligned} \quad (16)$$

其中

$$\mathbf{M} = (\mathbf{M}(t_0) \Delta t \quad \mathbf{M}(t_1) \Delta t \quad \dots \quad \mathbf{M}(t_{N-1}) \Delta t), \mathbf{c} = \boldsymbol{\eta} - \mathbf{x}_T$$

我们对线性系统(12) 作如下假设.

(A3) 线性系统(12) 完全能控, 即秩条件 $\text{rank}(\mathbf{A}, \mathbf{B}\mathbf{A}, \dots, \mathbf{B}^{n-1}\mathbf{A}) = n$ 成立.

在条件(A3) 下, 最优控制问题(14) 解存在唯一, 并且其离散优化问题(16) 中的矩阵 $\mathbf{S}, \mathbf{M}$ 满足假设条件(A1) – (A2).

引理 3^[14] 设 $\hat{\mathbf{u}}$ 与 $\tilde{\mathbf{u}}$ 分别是问题(14) 与(16) 的最优解, 记 $\tilde{\mathbf{u}}(t) = \sum_{i=1}^{N-1} \hat{\mathbf{u}}(t_i) \mathbf{X}_{[t_{i-1}, t_i]}$, 则

$$\|\tilde{\mathbf{u}}(t) - \hat{\mathbf{u}}(t)\|_{L^2} \leq C \Delta t$$

其中 C 为常数.

由引理 3, 求原范数最优控制问题(14) 可转化为求凸优化问题(16). 进一步, 在假设条件(A3) 下, 我们可用 ALM-BB 算法求解问题(16). 下面我们给出一个具体的算例, 并用它对原始的 ALM 与 ALM-BB 算法进行比较.

例 6 设 $T = 3, n = 2, m = 1$, 考虑如下控制系统:

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \text{ a. e. } t \in (0, 3) \\ \mathbf{x}(0) = \mathbf{x}_0, \mathbf{x}(3) = \mathbf{x}_T \end{cases}$$

及性能指标

$$J(\mathbf{u}) = \frac{1}{2} \int_0^T \|\mathbf{u}(t)\|^2 dt$$

其中 $\mathbf{A} = \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $\mathbf{x}_0 = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$ 是初始状态, $\mathbf{x}_T = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ 为终端状态约束. 我们直接利用算法 2 计算其对应的离散优化问题(16), 其中 $N=100$, 初始迭代点为 $\mathbf{u}_0 = (3, 3, \cdots, 3)$, $\boldsymbol{\lambda}_0 = \begin{pmatrix} 1 \\ 5 \end{pmatrix}$, 算法停止误差 $\epsilon = 10^{-6}$. 计算结果见图 1,2, 其中图 1 为两个分量 x_1 与 x_2 的最优状态图像, 图 2 为最优控制的图像.

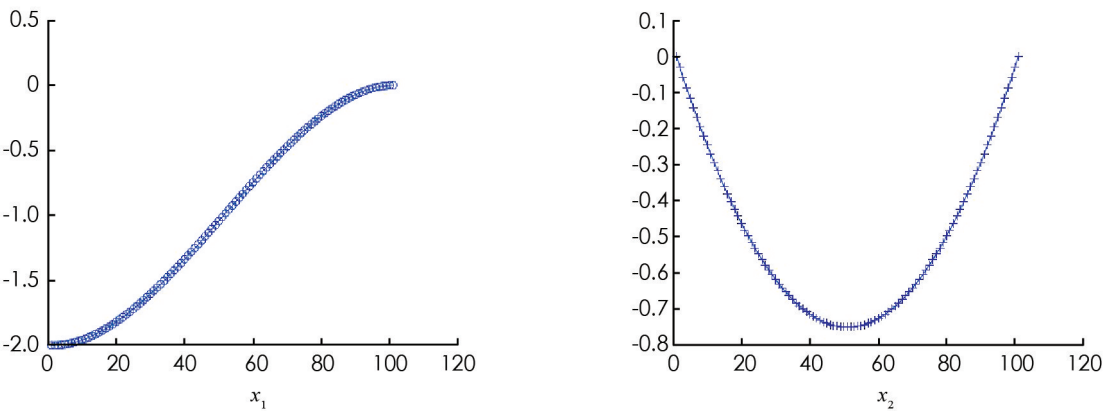


图 1 最优状态图

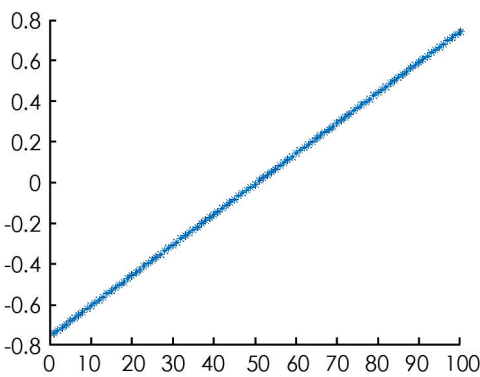


图 2 最优控制图

为了说明 ALM-BB 算法的有效性, 我们选择不同的罚参数和划分细度进行计算, 将 ALM-BB 和 ALM 算法的计算时间和迭代次数对比见表 1.

表 1 ALM-BB 与 ALM 数值实验结果对比

N	ρ	ALM-BB		ALM	
		时间/s	迭代次数/次	时间/s	迭代次数/次
100	100	0.444	4	0.957	41
100	500	0.262	3	1.076	41
100	5 000	0.619	3	0.712	41
100	10 000	0.588	3	0.937	41
200	100	0.271	3	1.019	41
200	500	0.355	3	1.028	41
200	5 000	0.623	3	0.665	41
200	10 000	0.627	3	0.940	41

由表 1 可以看出, 在划分细度与罚参数相同的情况下, ALB-BB 算法迭代步数更少, 收敛速度更快. 但是当罚参数选取过大时, ALM 与 ALM-BB 算法运行时间差别不大, 对罚参数的选取依然敏感.

4 总结

本文通过改变 ALM 算法中更新乘子这一迭代步的步长来提高算法的收敛速度. 范数最优控制问题本质上是无穷维的二次凸优化问题, 通过适当的离散格式可以将其转为有限维的二次凸优化问题. 通过求解范数最优控制问题的数值算例发现, ALM-BB 算法较 ALM 计算效率更高. 从数值实验看出 ALM-BB 算法的数值表现仍依赖于参数的选取. 此外, 本文的收敛性证明依赖于凸二次规划问题的具体结构以及子问题可精确求解, 对于更一般的非精确的 ALM-BB 算法的收敛性有待进一步研究.

参考文献:

- [1] COURANT R. Variational Methods for the Solution of Problems of Equilibrium and Vibrations [J]. Bulletin of the American Mathematical Society, 1943, 49(1): 1-23.
- [2] 刘浩洋, 户将, 李勇锋. 最优化: 建模、算法与理论 [M]. 北京: 高等教育出版社, 2020.
- [3] 宋菲, 吴泽忠. 外罚函数法与广义 Lagrange 乘子法的比较研究 [J]. 成都信息工程大学学报, 2017, 32(6): 667-674.
- [4] HESTENES M R. Multiplier and Gradient Methods [J]. Journal of Optimization Theory and Application, 1969, 4(5): 303-320.
- [5] POWELL M J D. A Method for Nonlinear Constraints in Minimization Problems [M]. New York: Academic Press, 1969: 283-298.
- [6] NOCEDAL J, WRIGHT S J. Numerical Optimization [M]. New York: Springer-Verlag, 1999.
- [7] NAGURNEY A, RAMANUJAM P. Transportation Network Policy Modeling with Goal Targets and Generalized Penalty Functions [J]. Transportation Science, 1996, 30(1): 3-13.
- [8] BARZILAI J, BORWEIN J M. Two-Point Step Size Gradient Methods [J]. IMA Journal of Numerical Analysis, 1988, 8(1): 141-148.
- [9] ZAREPISHEH M, XING L, YE Y Y. A Computation Study on an Integrated Alternating Direction Method of Multipliers for Large Scale Optimization [J]. Optimization Letters, 2018, 12(1): 3-15.
- [10] GABAY D. A Dual Algorithm for the Solution of Nonlinear Variational Problems via Finite Element Approximation [J]. Computers & Mathematics with Applications, 1976, 2(1): 17-40.
- [11] GLOWINSKI R, MARROCO A. Sur L'approximation, Par Éléments Finis D'ordre Un, et La Résolution, Par Pénalisation-Dualité D'une Classe de Problèmes de Dirichlet Non Linéaires [J]. Revue Française d'Automatique, Informatique, Recherche Opérationnelle Analyse Numérique, 1975, 9(2): 41-76.
- [12] 杜学武. 求解约束优化问题的增广拉格朗日函数法 [D]. 上海: 上海大学, 2005.
- [13] DAI Y H, LIAO L Z. R-Linear Convergence of the Barzilai and Borwein Gradient Method [J]. IMA Journal of Numerical Analysis, 2002, 22(1): 1-10.
- [14] DONTCHEV A L. Error Estimates for a Discrete Approximation to Constrained Control Problems [J]. SIAM Journal on Numerical Analysis, 1981, 18(3): 500-514.
- [15] 雍炯敏, 楼红卫. 最优控制理论简明教程 [M]. 北京: 高等教育出版社, 2006.

责任编辑 张枸