Advances in Applied Mathematics 应用数学进展, 2025, 14(4), 33-43
Published Online April 2025 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2025.144137

Hans汉斯

基于ALM的非精确加速算法

黄　芳

浙江师范大学数学科学学院，浙江　金华

收稿日期：2025年3月1日；录用日期：2025年3月25日；发布日期：2025年4月2日

摘　要

增广拉格朗日乘子法为经典有效的解决线性等式约束凸优化问题的一阶优化方法，算法通过原变量与对偶变量的交替迭代更新收敛至最优点。然而，子问题中原变量的更新在实际应用中往往无法精确求解。本文基于增广拉格朗日乘子法、对偶优化以及Nesterov加速技巧，提出一种非精确求解的增广拉格朗日乘子法，利用KKT条件从对偶残差的角度分析并从理论上证明该算法的收敛速率可达到$O\left(1/k^2\right)$。

关键词

增广拉格朗日乘子法，Nesterov加速，对偶理论

Inexact Acceleration Algorithm Based on ALM

Fang Huang

School of Mathematical Sciences, Zhejiang Normal University, Jinhua Zhejiang

Received: Mar. 1st, 2025; accepted: Mar. 25th, 2025; published: Apr. 2nd, 2025

Abstract

The Augmented Lagrangian Method is a classical and effective first-order optimization technique for solving convex optimization problems with linear equality constraints. The algorithm converges to the optimal solution through alternating iterative updates between the primal and dual variables. However, in practical applications, the update of the primal variables in the subproblem is often not solved exactly. In this paper, based on the Augmented Lagrangian Method, dual optimization, and Nesterov's acceleration technique, we propose an inexact solution version of the Augmented Lagrangian Method. By leveraging the Karush-Kuhn-Tucker (KKT) conditions, we analyze and prove that the convergence rate of the proposed algorithm can achieve a rate of $O\left(1/k^2\right)$.

文章引用: 黄芳. 基于 ALM 的非精确加速算法[J]. 应用数学进展, 2025, 14(4): 33-43.
DOI: 10.12677/aam.2025.144137

Keywords

ALM, Nesterov Acceleration, Duality Theory

1. 引言

现实世界中大量问题可转化为优化问题，特别地，具有特殊结构的凸优化问题是现代优化和计算数学领域的基石之一。原因在于以下两点：其一，凸优化问题具有良好的性质，如：局部最优解必然是全局最优解，并且具有丰富的理论背景，如最优性条件、对偶理论、拉格朗日乘数法等，这些理论为问题的分析、算法的设计及其性质提供了深入的理解；其二，现代科学和工程上的大量问题本质为求解凸优化问题，如信号处理与图像处理[1]-[4]、网络优化、控制理论、经济学与金融、机器学习与数据科学[5] [6]。为此，对于凸优化问题的研究显得尤为重要，特别是模型的求解以及加速算法成为了科学研究的重点与难点。

对于无约束光滑凸优化问题

$$\min_{x\in\mathcal{X}} f(x),$$

20 世纪 80 年代，Nesterov 在其开创性论文[7]中首次提出了仅使用一阶梯度信息且收敛速率可达到 $O(1/k^2)$ 的加速算法，其迭代格式如下：

$$\begin{cases} x_k = y_k - s\nabla f(y_k), \\ t_{k+1} = \dfrac{1+\sqrt{1+4t_k^2}}{2}, \\ y_{k+1} = x_k + \dfrac{t_k-1}{t_{k+1}}(x_k - x_{k-1}). \end{cases}$$

相较于基于当前位置的梯度来更新下一步位置的传统的梯度下降算法，Nesterov 加速梯度方法的特殊之处在于，它先外推到一个预测位置，然后基于该预测位置的梯度信息进行调整，为此该加速技术也被称为 Nesterov 外推法。Nesterov 加速算法提出时并没有受到太大的关注，原因在于当时的优化问题规模并不是很大，二阶优化方法可以非常有效地处理实际问题。随着越来越多领域优化问题规模的上升，Beck 和 Teboulle 于 2008 年给出了 Nesterov 在 1983 年提出的针对两个函数的复合优化问题

$$\min_{x\in\mathbb{R}^n} \psi(x) = f(x) + g(x),$$

的加速算法版本——FISTA [8]，一阶加速梯度算法开始广泛应用。FISTA 迭代格式如下：

$$\begin{cases} x_k = \mathrm{Prox}_{t_k g}\left(y_k - t_k\nabla f(y_k)\right), \\ t_{k+1} = \dfrac{1+\sqrt{1+4t_k^2}}{2}, \\ y_{k+1} = x_k + \dfrac{t_k-1}{t_{k+1}}(x_k - x_{k-1}), \end{cases}$$

其中，$f(x), g(x)$ 为凸函数，且其中至少有一个是光滑的。FISTA 核心思想为，第一步沿着前两步的计算方向计算一个新点，第二步在该新点处做一步近似点梯度迭代。此后，基于 Nesterov 加速技巧与 FISTA 思想受到了大量的关注与研究应用。

对于线性等式约束优化问题

$$\min_x f(x),$$
$$\text{s.t. } Ax = b,$$

其中 $f(x)$ 为凸集 $\mathcal{X}$ 上的凸函数，$A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$，经典而有效的求解方法为增广拉格朗日法(ALM, Augmented Lagrangian Method)，其迭代格式如下：

$$\begin{cases} x^{k+1} = \arg\min_x \left\{ L_\beta\left(x, \lambda^k\right), x \in \mathcal{X} \right\}, \\ \lambda^{k+1} = \lambda^k - \beta\left(Ax^{k+1} - b\right), \end{cases}$$

其中，增广拉格朗日函数

$$L_\beta(x, \lambda) = f(x) - \langle \lambda, Ax - b \rangle + \frac{\beta}{2}\|Ax - b\|^2, \quad \lambda \in \mathbb{R}^m.$$

ALM 最初由 Hestenes [9]和 Powell [10]提出的(它最初被称为[9] [10]中的乘子法；见[11]-[13])。相较于二次罚函数法(Quadratic Penalty Method)，ALM 通过引入对偶乘子，克服了原罚函数法中存在的病态问题。ALM 不仅在理论上有完整的收敛保证，并且在实际应用中具有高效的性能，为此 ALM 成为科研与工程应用的基础且成熟的有效算法。关于 ALM 算法的研究有大量文献，这里仅列取部分，见[8] [9] [11]。

随着优化问题规模的上升，如何基于 ALM 设计加速算法受到了大量科研人员的关注。当 $f(x)$ 可微时，He [14]结合 Nesterov 加速技巧[7]，提出收敛速度为 $O\left(1/k^2\right)$ 的加速 ALM，迭代格式如下：

$$\begin{cases} x^{k+1} = \arg\min_{x\in\mathcal{X}} \left\{ f(x) - \left\langle \lambda^k, Ax - b \right\rangle + \frac{1}{2}\|Ax - b\|_{H_k}^2 \right\}, \\ \tilde{\lambda}^k = \lambda^k - H_k\left(Ax^{k+1} - b\right), t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}, \\ \lambda^{k+1} = \tilde{\lambda}^k + \frac{t_k - 1}{t_{k+1}}\left(\tilde{\lambda}^k - \tilde{\lambda}^{k-1}\right) + \frac{t_k}{t_{k+1}}\left(\tilde{\lambda}^k - \lambda^k\right), \end{cases}$$

其中 $k \ge 0$，H 是任意的非单调递增对称正定矩阵。在此基础上，[15]给出 $f(x)$ 仅为凸函数时，收敛速率为 $O\left(1/k^2\right)$ 的加速 ALM。

在实际应用中，ALM 中的子问题 x^{k+1} 精确求解往往较为复杂，为此，是否可以在子问题非精确求解情况下设计算法使其仍可达到 $O\left(1/k^2\right)$ 的收敛速率？本文基于 ALM 与 Nesterov 加速技巧提出一种非精确的加速 ALM，并且从理论上给出收敛速率可达到 $O\left(1/k^2\right)$ 的证明。

2. 预备知识

为了方便算法的收敛性说明，本章对符号进行统一的说明并给出预备知识。

对任意常数 $p \in [0, \infty)$ 和向量 $x = (x_1, x_2, \cdots, x_n)^{\mathrm{T}} \in \mathbb{R}^n$，向量 x 的 l^p-范数定义为 $\|x\|_p := \left(\sum_{i=1}^n \left|x_i^p\right|\right)^{1/p}$。本文所述 $\|\cdot\|$ 均表示 l^2-范数。对任意向量 $x, y \in \mathbb{R}^n$，我们用 $x^{\mathrm{T}} y := \sum_{i=1}^n x_i y_i$ 表示 $\mathbb{R}^n$ 空间中的标准内积。

定义 2.1. 设函数 f 为适当函数，如果 ***dom*** f 是凸集，且

$$f\left(\theta x+(1-\theta)y\right)\le\theta f(x)+(1-\theta)f(y),$$

对所有 $x,y\in\boldsymbol{dom}\ f,0\le\theta\le1$ 都成立，则称 f 是凸函数。

定义 2.2. 凸函数 $f:\mathbb{R}^n\to\mathbb{R}$，其次微分为

$$\partial f(x)=\left\{s:f(y)-f(x)\ge s^{\mathrm{T}}(y-x),\forall y\in\mathbb{R}^n\right\}.$$

定义 2.3. 函数 f 的共轭函数 $f^*:\mathbb{R}^n\to\mathbb{R}$ 定义如下

$$f^*(x)=\max_{p\in\mathbb{R}^n}\left\{x^{\mathrm{T}}p-f(p)\right\}.$$

性质 1. 共轭函数 f^* 的次微分性质

$$y\in\partial f(x)\Leftrightarrow x\in\partial f^*(y).$$

定义 2.4. 函数 f 是一个参数为 σ_f 的强凸函数，当且仅当存在一个常数 $\sigma_f>0$，使得函数 $f(x)-\frac{\sigma_f}{2}\|x\|^2$ 是凸函数，且对每一个 x 和 y 满足以下不等式：

$$\langle p-q,x-y\rangle\ge\sigma_f\|x-y\|^2,\quad p\in\partial f(x),q\in\partial f(y).$$

引理 2.1 若 f 是凸性参数为 σ_f 的强凸函数，则共轭函数 f^* 可微且 ∇f^* 是利普希茨连续函数，其利普西茨常数 $L\left(\nabla f^*\right)=\frac{1}{\sigma_f}$，即满足下列不等式：

$$\nabla f^*(x)-\nabla f^*(y)\le\frac{1}{\sigma_f}\|x-y\|.$$

3. 非精确加速 ALM

对于线性等式约束优化问题

$$\begin{aligned}&\min_x f(x),\\&\text{s.t. } Ax=b,\end{aligned}\tag{1}$$

其中 $f(x)$ 为凸集 $\mathcal{X}$ 上的凸函数，$A\in\mathbb{R}^{m\times n},b\in\mathbb{R}^m$。本章具体展示以上优化问题的非精确加速 ALM，在此之前，先给出以下假设：

假设 1. 优化问题(1)中的函数 $f(x)$ 是 σ_f -强凸函数。

基于以上假设，下面给出优化问题(1)的非精确加速 ALM，具体迭代格式如下：给定初值 $\hat{\lambda}_1=\lambda_0$，并作如下更新，

$$(\text{I-AALM})\begin{cases}x_k\approx\arg\min_x\left\{f(x)-\left(\hat{\lambda}_k\right)^{\mathrm{T}}(Ax-b)+\frac{\beta}{2}\|Ax-b\|^2\right\},\\\lambda_k=\hat{\lambda}_k-\beta(Ax_k-b),\\\hat{\lambda}_{k+1}=\lambda_k+(1-\theta_{k-1})\frac{\theta_k}{\theta_{k-1}}(\lambda_k-\lambda_{k-1}),\end{cases}\tag{2}$$

其中 $\theta_k=\frac{2}{k+2},k\ge0$。

由算法 I-AALM 中子问题 x_k 最优性条件可知

$$0 \in \partial f\left(x_k\right) - A^{\mathrm{T}}\hat{\lambda}_k + \beta A^{\mathrm{T}}\left(Ax_k - b\right),$$

结合算法 I-AALM 中 λ_k 的更新方式，可得

$$A^{\mathrm{T}}\lambda_k \in \partial f\left(x_k\right).$$

记误差向量 $\delta_k = f'\left(x_k\right) - A^{\mathrm{T}}\lambda_k$，其中 $f'\left(x_k\right) \in \partial f\left(x_k\right)$，所以

$$\begin{cases} \delta_k = 0,\ x_k \text{ 是精确解}, \\ \delta_k \neq 0,\ x_k \text{ 是非精确解}. \end{cases}$$

由 δ_k 的记法和性质 1 可得

$$A^{\mathrm{T}}\lambda_k + \delta_k \in \partial f\left(x_k\right) \Rightarrow x_k \in \nabla f^*\left(A^{\mathrm{T}}\lambda_k + \delta_k\right) \Rightarrow Ax_k \in A\nabla f^*\left(A^{\mathrm{T}}\lambda_k + \delta_k\right).$$

记误差向量 $\eta_k = A\nabla f^*\left(A^{\mathrm{T}}\lambda_k\right) - A\nabla f^*\left(A^{\mathrm{T}}\lambda_k + \delta_k\right)$。若 f 是凸性参数为 σ_f 的强凸函数，易知 ∇f^* 利普西茨连续，所以误差向量的上界易得

$$\left\|\eta_k\right\| = \left\|A\nabla f^*\left(A^{\mathrm{T}}\lambda_k\right) - A\nabla f^*\left(A^{\mathrm{T}}\lambda_k + \delta_k\right)\right\| \leq \frac{1}{\sigma_f}\sqrt{\rho\left(A^{\mathrm{T}}A\right)}\left\|\delta_k\right\|.$$

为此，子问题 x_k 求解的停止条件为：

$$\delta_k = f'\left(x_k\right) - A^{\mathrm{T}}\lambda_k, \left\|\delta_k\right\| \leq \frac{\sigma_f\theta_{k-1}}{\sqrt{\rho\left(A^{\mathrm{T}}A\right)}}\epsilon_k, \tag{3}$$

其中 $\epsilon_k \geq 0$ 且满足 $\sum_{k=1}^{\infty}\epsilon_k < \infty, \epsilon_{k+1} \leq \epsilon_k$。

4. 收敛性分析

本章给出非精确加速 ALM 的收敛分析并说明其收敛速率为 $O\left(1/k^2\right)$。问题(1)的对偶函数为

$$\begin{aligned} D(\lambda) &= \min_x\left\{f(x) - \lambda^{\mathrm{T}}(Ax - b)\right\} \\ &= -\max_x\left\{\lambda^{\mathrm{T}}Ax - f(x)\right\} + \lambda^{\mathrm{T}}b, \\ &= -f^*\left(A^{\mathrm{T}}\lambda\right) + \lambda^{\mathrm{T}}b. \end{aligned}$$

不难得到问题(1)的对偶问题

$$\max_\lambda D(\lambda) = \max_\lambda\left\{-f^*\left(A^{\mathrm{T}}\lambda\right) + \lambda^{\mathrm{T}}b\right\}. \tag{4}$$

因此，对原问题(1)的研究可转换为对对偶问题(4)的研究，本章以下内容将围绕问题(4)及假设 1 进行讨论。在此之前，我们先给出以下引理，这对本文算法的收敛性分析有着至关重要的作用。

引理 4.1. 算法 I-AALM 中参数 θ_k 满足 $\frac{1-\theta_{k+1}}{\theta_{k+1}^2} \leq \frac{1}{\theta_k^2}, \forall k \geq 0$。

证明. 数学归纳法易证。

引理 4.2. 记 $s_k = \frac{1}{\theta_{k-1}}\lambda_k - \left(\frac{1}{\theta_{k-1}} - 1\right)\lambda_{k-1} - \lambda^*$，则 $s_{k+1} = s_k + \frac{1}{\theta_k}\left(\lambda_{k+1} - \hat{\lambda}_{k+1}\right)$.

证明. 由 s_k 的记法，有

$$
\begin{aligned}
s_{k+1} &= \frac{1}{\theta_k}\lambda_{k+1} - \left(\frac{1}{\theta_k} - 1\right)\lambda_k - \lambda^* \\
&= s_k + \frac{1}{\theta_k}\lambda_{k+1} - \frac{1}{\theta_k}\lambda_k + \left(1 - \frac{1}{\theta_{k-1}}\right)\left(\lambda_k - \lambda_{k-1}\right) \\
&= s_k + \frac{1}{\theta_k}\left(\lambda_{k+1} - \lambda_k + \left(1 - \frac{1}{\theta_{k-1}}\right)\theta_k\left(\lambda_k - \lambda_{k-1}\right)\right) \\
&= s_k + \frac{1}{\theta_k}\left(\lambda_{k+1} - \hat{\lambda}_{k+1}\right).
\end{aligned}
$$

引理 4.3. 设$\left\{\lambda_{k+1}, \hat{\lambda}_{k+1}\right\}$为 I-AALM 生成的点列，对任何的$\lambda \in R^m, k \ge 0$，有

$$
D\left(\lambda_{k+1}\right) - D\left(\lambda\right) \ge \frac{1}{\beta}\left(\lambda - \hat{\lambda}_{k+1}\right)^{\mathrm{T}}\left(\hat{\lambda}_{k+1} - \lambda_{k+1}\right) + \frac{1}{2\beta}\left\|\lambda_{k+1} - \hat{\lambda}_{k+1}\right\|^2 + \left(\lambda - \lambda_{k+1}\right)^{\mathrm{T}}\eta_{k+1}, \tag{5}
$$

其中$\eta_{k+1} = A\nabla f^*\left(A^{\mathrm{T}}\lambda_{k+1}\right) - A\nabla f^*\left(A^{\mathrm{T}}\lambda_{k+1} + \delta_{k+1}\right)$。

证明. 因为$f\left(x\right)$是σ_f-强凸函数，所以其共轭函数f^*是凸函数，于是

$$
f^*\left(A^{\mathrm{T}}\lambda\right) - f^*\left(A^{\mathrm{T}}\lambda_{k+1}\right) \ge \left(A^{\mathrm{T}}\left(\lambda - \lambda_{k+1}\right)\right)^{\mathrm{T}}\nabla f^*\left(A^{\mathrm{T}}\lambda_{k+1}\right) = \left(\lambda - \lambda_{k+1}\right)^{\mathrm{T}}A\nabla f^*\left(A^{\mathrm{T}}\lambda_{k+1}\right),
$$

进一步根据引理 4.3.中η_{k+1}的记法可得

$$
\begin{aligned}
& D\left(\lambda_{k+1}\right) - D\left(\lambda\right) \\
&= f^*\left(A^{\mathrm{T}}\lambda\right) - f^*\left(A^{\mathrm{T}}\lambda_{k+1}\right) + \left(\lambda_{k+1} - \lambda\right)^{\mathrm{T}}b \\
&\ge \left(\lambda - \lambda_{k+1}\right)^{\mathrm{T}}A\nabla f^*\left(A^{\mathrm{T}}\lambda_{k+1}\right) + \left(\lambda_{k+1} - \lambda\right)^{\mathrm{T}}b \\
&= \left(\lambda - \lambda_{k+1}\right)^{\mathrm{T}}\eta_{k+1} + \left(\lambda - \lambda_{k+1}\right)^{\mathrm{T}}\left(-b\right) + \left(\lambda - \lambda_{k+1}\right)^{\mathrm{T}}A\nabla f^*\left(A^{\mathrm{T}}\lambda_{k+1} + \delta_{k+1}\right) \\
&= \left(\lambda - \lambda_{k+1}\right)^{\mathrm{T}}\eta_{k+1} + \left(\lambda - \lambda_{k+1}\right)^{\mathrm{T}}\left(Ax_{k+1} - b\right) \\
&= \frac{1}{\beta}\left(\lambda - \lambda_{k+1}\right)^{\mathrm{T}}\left(\hat{\lambda}_{k+1} - \lambda_{k+1}\right) + \left(\lambda - \lambda_{k+1}\right)^{\mathrm{T}}\eta_{k+1} \\
&= \frac{1}{\beta}\left(\lambda - \hat{\lambda}_{k+1}\right)^{\mathrm{T}}\left(\hat{\lambda}_{k+1} - \lambda_{k+1}\right) + \frac{1}{\beta}\left\|\hat{\lambda}_{k+1} - \lambda_{k+1}\right\|^2 + \left(\lambda - \lambda_{k+1}\right)^{\mathrm{T}}\eta_{k+1} \\
&\ge \frac{1}{\beta}\left(\lambda - \hat{\lambda}_{k+1}\right)^{\mathrm{T}}\left(\hat{\lambda}_{k+1} - \lambda_{k+1}\right) + \frac{1}{2\beta}\left\|\hat{\lambda}_{k+1} - \lambda_{k+1}\right\|^2 + \left(\lambda - \lambda_{k+1}\right)^{\mathrm{T}}\eta_{k+1},
\end{aligned}
$$

其中倒数第二个等号成立是根据$\left\{\lambda_k\right\}$的更新方式。

引理 4.4. 设$\left\{\lambda_k\right\}$为 I-AALM 生成的点列，记$h_k' = D\left(\lambda^*\right) - D\left(\lambda_k\right)$，那么

$$
\frac{1}{\theta_{k-1}^2}h_k' - \frac{1}{\theta_k^2}h_{k+1}' \ge \frac{\left\|s_{k+1}\right\|^2}{2\beta} - \frac{\left\|s_k\right\|^2}{2\beta} - \frac{1}{\theta_k}\left(s_{k+1}\right)^{\mathrm{T}}\eta_{k+1}. \tag{6}
$$

证明. 由$h_k' = D\left(\lambda^*\right) - D\left(\lambda_k\right)$，易得到$h_k' \ge 0$，引理 4.3.中对不等式(5)分别令$\lambda = \lambda_k, \lambda = \lambda^*$可得

$$
h_k' - h_{k+1}' \ge \frac{1}{2\beta}\left\|\hat{\lambda}_{k+1} - \lambda_{k+1}\right\|^2 + \frac{1}{\beta}\left(\lambda_k - \hat{\lambda}_{k+1}\right)^{\mathrm{T}}\left(\hat{\lambda}_{k+1} - \lambda_{k+1}\right) + \left(\lambda_k - \lambda_{k+1}\right)^{\mathrm{T}}\eta_{k+1}, \tag{7}
$$

$$
-h_{k+1}' \ge \frac{1}{2\beta}\left\|\hat{\lambda}_{k+1} - \lambda_{k+1}\right\|^2 + \frac{1}{\beta}\left(\lambda^* - \hat{\lambda}_{k+1}\right)^{\mathrm{T}}\left(\hat{\lambda}_{k+1} - \lambda_{k+1}\right) + \left(\lambda^* - \lambda_{k+1}\right)^{\mathrm{T}}\eta_{k+1}. \tag{8}
$$

对以上两个不等式作如下处理：(7)式乘以$\left(\frac{1}{\theta_k} - 1\right)$加到(8)式，那么

$$\left(\frac{1}{\theta_k}-1\right)h_k'-\frac{1}{\theta_k}h_{k+1}'\geq\frac{1}{2\beta\theta_k}\left\|\hat{\lambda}_{k+1}-\lambda_{k+1}\right\|^2+\frac{1}{\beta}\left(\left(\frac{1}{\theta_k}-1\right)\lambda_k-\frac{1}{\theta_k}\hat{\lambda}_{k+1}+\lambda^*\right)^{\mathrm{T}}\left(\hat{\lambda}_{k+1}-\lambda_{k+1}\right)$$
$$+\left(\left(\frac{1}{\theta_k}-1\right)\lambda_k-\frac{1}{\theta_k}\hat{\lambda}_{k+1}+\lambda^*\right)^{\mathrm{T}}\eta_{k+1}$$

上述不等式两端同时与$\frac{1}{\theta_k}$相乘得

$$\frac{1-\theta_k}{\theta_k^2}h_k'-\frac{1}{\theta_k^2}h_{k+1}'\geq\frac{1}{2\beta}\left\|\frac{1}{\theta_k}\left(\hat{\lambda}_{k+1}-\lambda_{k+1}\right)\right\|^2+\frac{1}{\beta\theta_k}\left(\hat{\lambda}_{k+1}-\lambda_{k+1}\right)^{\mathrm{T}}\left(\left(\frac{1}{\theta_k}-1\right)\lambda_k-\frac{1}{\theta_k}\hat{\lambda}_{k+1}+\lambda^*\right)$$
$$+\frac{1}{\theta_k}\left(\left(\frac{1}{\theta_k}-1\right)\lambda_k-\frac{1}{\theta_k}\lambda_{k+1}+\lambda^*\right)^{\mathrm{T}}\eta_{k+1}.$$

由$\frac{1-\theta_{k+1}}{\theta_{k+1}^2}\leq\frac{1}{\theta_k^2},\forall k\geq 0$，上式进一步可化为

$$\frac{1}{\theta_{k-1}^2}h_k'-\frac{1}{\theta_k^2}h_{k+1}'\geq\frac{1}{2\beta}\left\|\frac{1}{\theta_k}\left(\hat{\lambda}_{k+1}-\lambda_{k+1}\right)\right\|^2+\frac{1}{\beta\theta_k}\left(\hat{\lambda}_{k+1}-\lambda_{k+1}\right)^{\mathrm{T}}\left(\left(\frac{1}{\theta_k}-1\right)\lambda_k-\frac{1}{\theta_k}\hat{\lambda}_{k+1}+\lambda^*\right)$$
$$+\frac{1}{\theta_k}\left(\left(\frac{1}{\theta_k}-1\right)\lambda_k-\frac{1}{\theta_k}\lambda_{k+1}+\lambda^*\right)^{\mathrm{T}}\eta_{k+1}.$$

取$\boldsymbol{a}=\frac{1}{\theta_k}\left(\hat{\lambda}_{k+1}-\lambda_{k+1}\right),\boldsymbol{b}=\left(\frac{1}{\theta_k}-1\right)\lambda_k-\frac{1}{\theta_k}\hat{\lambda}_{k+1}+\lambda^*$，并利用等式关系

$$\|\boldsymbol{a}\|+2\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}=\|\boldsymbol{a}+\boldsymbol{b}\|^2-\|\boldsymbol{b}\|^2,$$

以上不等式可进一步转化为

$$\begin{aligned}
&\frac{1}{\theta_{k-1}^2}h_k'-\frac{1}{\theta_k^2}h_{k+1}'\\
&\geq\frac{1}{2\beta}\left\|\frac{1}{\theta_k}\left(\hat{\lambda}_{k+1}-\lambda_{k+1}\right)+\left(\frac{1}{\theta_k}-1\right)\lambda_k-\frac{1}{\theta_k}\hat{\lambda}_{k+1}+\lambda^*\right\|^2-\frac{1}{2\beta}\left\|\left(\frac{1}{\theta_k}-1\right)\lambda_k-\frac{1}{\theta_k}\hat{\lambda}_{k+1}+\lambda^*\right\|^2\\
&+\frac{1}{\theta_k}\left(\left(\frac{1}{\theta_k}-1\right)\lambda_k-\frac{1}{\theta_k}\lambda_{k+1}+\lambda^*\right)^{\mathrm{T}}\eta_{k+1}\\
&=\frac{1}{2\beta}\left\|-\frac{1}{\theta_k}\lambda_{k+1}+\left(\frac{1}{\theta_k}-1\right)\lambda_k+\lambda^*\right\|^2-\frac{1}{2\beta}\left\|\left(\frac{1}{\theta_k}-1\right)\lambda_k-\frac{1}{\theta_k}\hat{\lambda}_{k+1}+\lambda^*\right\|^2\\
&+\frac{1}{\theta_k}\left(\left(\frac{1}{\theta_k}-1\right)\lambda_k-\frac{1}{\theta_k}\lambda_{k+1}+\lambda^*\right)^{\mathrm{T}}\eta_{k+1}
\end{aligned}$$

由引理 4.2 中 s_k 的记法，上式转化为

$$\frac{1}{\theta_{k-1}^2}h_k'-\frac{1}{\theta_k^2}h_{k+1}'\geq\frac{1}{2\beta}\|s_{k+1}\|^2-\frac{1}{2\beta}\left\|\left(\frac{1}{\theta_k}-1\right)\lambda_k-\frac{1}{\theta_k}\hat{\lambda}_{k+1}+\lambda^*\right\|^2-\frac{1}{\theta_k}\left(s_{k+1}\right)^{\mathrm{T}}\eta_{k+1}. \tag{9}$$

为满足(6)的形式，仅需将不等式(9)右端做如下处理

$$\left(1-\frac{1}{\theta_k}\right)\lambda_k+\frac{1}{\theta_k}\hat{\lambda}_{k+1}-\lambda^*=\frac{1}{\theta_{k-1}}\lambda_k-\left(\frac{1}{\theta_{k-1}}-1\right)\lambda_{k-1}-\lambda^*,$$

化简整理后有

$$\hat{\lambda}_{k+1}=\lambda_k+\left(1-\theta_{k-1}\right)\frac{\theta_k}{\theta_{k-1}}\left(\lambda_k-\lambda_{k-1}\right),$$

所以 $\frac{1}{\theta_{k-1}^2}h_k'-\frac{1}{\theta_k^2}h_{k+1}'\geq\frac{\|s_{k+1}\|^2}{2\beta}-\frac{\|s_k\|^2}{2\beta}-\frac{1}{\theta_k}\left(s_{k+1}\right)^{\mathrm{T}}\eta_{k+1}$ 得证。

下面展示我们的主要收敛性定理，定理直观地解释了算法的收敛速度可达到 $O\left(1/k^2\right)$。

定理 4.1. 设 $\left\{\lambda_{k+1},\hat{\lambda}_{k+1}\right\}$ 为 I-AALM 生成的点列，那么

$$h_k'\leq\frac{4}{(k+1)^2}\left(\left(\sqrt{2\beta}\bar{\epsilon}_k+\frac{1}{\sqrt{2\beta}}\left\|\lambda^*-\hat{\lambda}_1\right\|^2\right)^2+2\tilde{\epsilon}_k\right)$$

其中 $h_k'=D\left(\lambda^*\right)-D\left(\lambda_k\right)$，$\bar{\epsilon}_k=\sum_{j=1}^{k}\epsilon_j$，$\tilde{\epsilon}_k=2\beta\sum_{j=1}^{k}\left(\epsilon_j\right)^2$。

证明. 为了方便证明，我们设置以下记号：

$$h_k=\frac{1}{\theta_{k-1}^2}h_k',\ p_k=\frac{\|s_k\|^2}{2\beta},\tag{10}$$

由引理 4.4 中 h_k' 的记法可知 $h_k'\geq 0$。引理 4.3.中，令 $\lambda=\lambda^*,k=0$ 得

$$\begin{aligned}-h_1=-h_1'&\geq\frac{1}{\beta}\left(\lambda^*-\hat{\lambda}_1\right)^{\mathrm{T}}\left(\hat{\lambda}_1-\lambda_1\right)+\frac{1}{2\beta}\left\|\lambda_1-\hat{\lambda}_1\right\|^2+\left(\lambda^*-\lambda_1\right)^{\mathrm{T}}\eta_1\\&=\frac{1}{2\beta}\left\|\lambda_1-\lambda^*\right\|^2-\frac{1}{2\beta}\left\|\lambda^*-\hat{\lambda}_1\right\|^2+\left(\lambda^*-\lambda_1\right)^{\mathrm{T}}\eta_1\\&=p_1-\frac{1}{2\beta}\left\|\lambda^*-\hat{\lambda}_1\right\|^2-\left(s_1\right)^{\mathrm{T}}\eta_1.\end{aligned}$$

其中第二个等号成立是因为等式关系

$$\|\boldsymbol{a}\|+2\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}=\|\boldsymbol{a}+\boldsymbol{b}\|^2-\|\boldsymbol{b}\|^2,$$

并且根据 x_k 的停止条件以及(10)中 p_k 的记法易知

$$\left(s_k\right)^{\mathrm{T}}\eta_k\leq\|s_k\|\cdot\|\eta_k\|\leq\|s_k\|\left(\frac{1}{\sigma_f}\sqrt{\rho\left(A^{\mathrm{T}}A\right)}\|\delta_k\|\right)\leq\|s_k\|\theta_{k-1}\epsilon_k=\sqrt{2\beta p_k}\theta_{k-1}\epsilon_k,$$

所以可以得到

$$\begin{aligned}p_1+h_1&\leq\frac{1}{2\beta}\left\|\lambda^*-\hat{\lambda}_1\right\|^2+\left(s_1\right)^{\mathrm{T}}\eta_1\leq\frac{1}{2\beta}\left\|\lambda^*-\hat{\lambda}_1\right\|^2+\sqrt{2\beta p_1}\theta_0\epsilon_1\\&=\frac{1}{2\beta}\left\|\lambda^*-\hat{\lambda}_1\right\|^2+\sqrt{2\beta p_1}\epsilon_1.\end{aligned}\tag{11}$$

由引理 4.4 知

$$h_k-h_{k+1}\geq p_{k+1}-p_k-\frac{1}{\theta_k}\left(s_{k+1}\right)^{\mathrm{T}}\eta_{k+1}$$

即 $h_k - h_{k+1} \ge p_{k+1} - p_k - \frac{1}{\theta_k}\sqrt{2\beta p_{k+1}}\theta_k \epsilon_{k+1}$。

两边求和，我们有

$$h_1 - h_k \ge p_k - p_1 - \sum_{j=1}^{k-1}\sqrt{2\beta p_{j+1}}\epsilon_{j+1}.$$

所以(11)经过变形可得

$$\frac{1}{2\beta}\left\|\lambda^* - \hat{\lambda}_1\right\|^2 \ge p_1 + h_1 - \sqrt{2\beta p_1}\epsilon_1 \ge h_k + p_k - q_k, \tag{12}$$

其中 $q_k := \sum_{j=1}^{k}\sqrt{2\beta p_j}\epsilon_j \ge 0$，由 ϵ_k 及 p_k 的记法，见(3)和(10)，可知 $q_k \le q_{k+1}$，所以

$$q_k = q_{k-1} + \epsilon_k\sqrt{2\beta p_k} \le q_{k-1} + \epsilon_k\sqrt{2\beta}\sqrt{q_k + \frac{1}{2\beta}\left\|\lambda^* - \hat{\lambda}_1\right\|^2}. \tag{13}$$

上式等价于

$$\left(q_k + \frac{1}{2\beta}\left\|\lambda^* - \hat{\lambda}_1\right\|^2\right) - \epsilon_k\sqrt{2\beta}\sqrt{q_k + \frac{1}{2\beta}\left\|\lambda^* - \hat{\lambda}_1\right\|^2} - \left(q_{k-1} + \frac{1}{2\beta}\left\|\lambda^* - \hat{\lambda}_1\right\|^2\right) \le 0,$$

解得

$$\begin{aligned}
\sqrt{q_k + \frac{1}{2\beta}\left\|\lambda^* - \hat{\lambda}_1\right\|^2} &\le \frac{1}{2}\left(\epsilon_k\sqrt{2\beta} + \sqrt{2\beta\epsilon_k^2 + 4\left(q_{k-1} + \frac{1}{2\beta}\left\|\lambda^* - \hat{\lambda}_1\right\|^2\right)}\right) \\
&\le \frac{1}{2}\left(\epsilon_k\sqrt{2\beta} + \epsilon_k\sqrt{2\beta} + 2\sqrt{q_{k-1} + \frac{1}{2\beta}\left\|\lambda^* - \hat{\lambda}_1\right\|^2}\right) \\
&= \epsilon_k\sqrt{2\beta} + \sqrt{q_{k-1} + \frac{1}{2\beta}\left\|\lambda^* - \hat{\lambda}_1\right\|^2} \\
&\le \epsilon_k\sqrt{2\beta} + \sqrt{q_{k-1}} + \frac{1}{\sqrt{2\beta}}\left\|\lambda^* - \hat{\lambda}_1\right\|
\end{aligned}$$

其中第二个和最后一个不等号成立均因为 $\sqrt{a^2+b^2} \le a+b, (a \ge 0, b \ge 0)$，所以由(13)知

$$\begin{aligned}
q_k &\le q_{k-1} + \epsilon_k\sqrt{2\beta}\left(\epsilon_k\sqrt{2\beta} + \sqrt{q_{k-1}} + \frac{1}{\sqrt{2\beta}}\left\|\lambda^* - \hat{\lambda}_1\right\|\right) \\
&= q_{k-1} + \epsilon_k^2 2\beta + \epsilon_k\left\|\lambda^* - \hat{\lambda}_1\right\| + \epsilon_k\sqrt{2\beta}\sqrt{q_{k-1}}.
\end{aligned}$$

对上式两边求和，有

$$q_k \le q_1 + 2\beta\sum_{j=2}^{k}\epsilon_j^2 + \left\|\lambda^* - \hat{\lambda}_1\right\|\sum_{j=2}^{k}\epsilon_j + \sqrt{2\beta}\sum_{j=2}^{k}\epsilon_j\sqrt{q_{j-1}}. \tag{14}$$

又由(11)及 $h_1 \ge 0$ 知

$$p_1 - \epsilon_1\sqrt{2\beta}\sqrt{p_1} - \frac{1}{2\beta}\left\|\lambda^* - \hat{\lambda}_1\right\|^2 \le 0,$$

同理可解得

$$\begin{aligned}\sqrt{p_1} &\le \frac{1}{2}\left(2\epsilon_1\sqrt{2\beta}+\sqrt{\epsilon_1^2 2\beta+4\left(\frac{1}{2\beta}\left\|\lambda^*-\hat{\lambda}_1\right\|^2\right)}\right)\\ &\le \frac{1}{2}\left(\epsilon_1\sqrt{2\beta}+\epsilon_1\sqrt{2\beta}+2\frac{1}{\sqrt{2\beta}}\left\|\lambda^*-\hat{\lambda}_1\right\|\right)\\ &= \epsilon_1\sqrt{2\beta}+\frac{1}{\sqrt{2\beta}}\left\|\lambda^*-\hat{\lambda}_1\right\|\end{aligned}$$

其中第一个不等号成立是因为$\sqrt{a^2+b^2}\le a+b,(a\ge 0,b\ge 0)$所以

$$q_1=\epsilon_1\sqrt{2\beta}\sqrt{p_1}\le\epsilon_1\sqrt{2\beta}\left(\epsilon_1\sqrt{2\beta}+\frac{1}{\sqrt{2\beta}}\left\|\lambda^*-\hat{\lambda}_1\right\|\right)=2\beta\epsilon_1^2+\epsilon_1\left\|\lambda^*-\hat{\lambda}_1\right\|.$$

把上述不等式代入(14)，得到

$$\begin{aligned}q_k &\le 2\beta\epsilon_1^2+\epsilon_1\left\|\lambda^*-\hat{\lambda}_1\right\|+2\beta\sum_{j=2}^{k}\epsilon_j^2+\left\|\lambda^*-\hat{\lambda}_1\right\|\sum_{j=2}^{k}\epsilon_j+\sqrt{2\beta}\sum_{j=2}^{k}\epsilon_j\sqrt{q_{j-1}}\\ &\le 2\beta\sum_{j=1}^{k}\epsilon_j^2+\left\|\lambda^*-\hat{\lambda}_1\right\|\sum_{j=1}^{k}\epsilon_j+\sqrt{2\beta q_k}\sum_{j=1}^{k}\epsilon_j\\ &= \tilde{\epsilon}_k+\left\|\lambda^*-\hat{\lambda}_1\right\|\bar{\epsilon}_k+\sqrt{2\beta q_k}\bar{\epsilon}_k,\end{aligned}$$

其中最后一个不等号成立是因为$q_k\le q_{k+1}$。所以上式转化为

$$q_k-\sqrt{2\beta q_k}\bar{\epsilon}_k-\left(\left\|\lambda^*-\hat{\lambda}_1\right\|\bar{\epsilon}_k+\tilde{\epsilon}_k\right)\le 0,$$

同理可解得

$$\begin{aligned}\sqrt{q_k} &\le \frac{1}{2}\left(\sqrt{2\beta}\bar{\epsilon}_k+\sqrt{2\beta\bar{\epsilon}_k^2+4\left(\left\|\lambda^*-\hat{\lambda}_1\right\|\bar{\epsilon}_k+\tilde{\epsilon}_k\right)}\right)\\ &\le \sqrt{\frac{1}{2}\left(2\beta\bar{\epsilon}_k^2+2\beta\bar{\epsilon}_k^2+4\left(\left\|\lambda^*-\hat{\lambda}_1\right\|\bar{\epsilon}_k+\tilde{\epsilon}_k\right)\right)}\\ &= \sqrt{2\beta\bar{\epsilon}_k^2+2\left(\left\|\lambda^*-\hat{\lambda}_1\right\|\bar{\epsilon}_k+\tilde{\epsilon}_k\right)},\end{aligned}$$

其中最后一个不等式成立是因为基本不等式$\frac{a+b}{2}\le\sqrt{\frac{a^2+b^2}{2}}$。结合上述不等式以及(12)和$p_k\ge 0$，

$$\begin{aligned}h_k &\le \frac{1}{2\beta}\left\|\lambda^*-\hat{\lambda}_1\right\|^2+q_k\le\frac{1}{2\beta}\left\|\lambda^*-\hat{\lambda}_1\right\|^2+2\beta\bar{\epsilon}_k^2+2\left(\left\|\lambda^*-\hat{\lambda}_1\right\|\bar{\epsilon}_k+\tilde{\epsilon}_k\right)\\ &= \left(\frac{1}{\sqrt{2\beta}}\left\|\lambda^*-\hat{\lambda}_1\right\|+\sqrt{2\beta}\bar{\epsilon}_k\right)^2+2\tilde{\epsilon}_k.\end{aligned}$$

5. 小结

ALM 是一类解决线性等式约束凸优化问题的有效方法，不仅在理论上有完整的收敛保证，并且在实际应用中具有高效的性能。然而经典的 ALM 中子问题通常情况下无法精确求解，且收敛速率为$O(1/k)$。为此，本文基于原优化模型的对偶问题，结合子问题的非精确解对应的停止条件与 Nesterov 加速技巧设计非精确加速 ALM 框架，并利用 KKT 条件从对偶残差的角度证明该算法收敛且收敛速率为$O(1/k^2)$。

参考文献

[1] Figueiredo, M.A.T. and Bioucas-Dias, J.M. (2010) Restoration of Poissonian Images Using Alternating Direction Optimization. *IEEE Transactions on Image Processing*, **19**, 3133-3145. https://doi.org/10.1109/tip.2010.2053941

[2] Yang, J. and Zhang, Y. (2011) Alternating Direction Algorithms for ℓ_1-Problems in Compressive Sensing. *SIAM Journal on Scientific Computing*, **33**, 250-278. https://doi.org/10.1137/090777761

[3] Yang, J., Zhang, Y. and Yin, W. (2010) A Fast Alternating Direction Method for TVL1-L2 Signal Reconstruction from Partial Fourier Data. *IEEE Journal of Selected Topics in Signal Processing*, **4**, 288-297. https://doi.org/10.1109/jstsp.2010.2042333

[4] Rudin, L.I., Osher, S. and Fatemi, E. (1992) Nonlinear Total Variation Based Noise Removal Algorithms. *Physica D*: *Nonlinear Phenomena*, **60**, 259-268. https://doi.org/10.1016/0167-2789(92)90242-f

[5] Liu, Z., Li, J., Li, G., Bai, J. and Liu, X. (2017) A New Model for Sparse and Low-Rank Matrix Decomposition. *Journal of Applied Analysis & Computation*, **7**, 600-616. https://doi.org/10.11948/2017037

[6] Bai, J., Hager, W.W. and Zhang, H. (2022) An Inexact Accelerated Stochastic ADMM for Separable Convex Optimization. *Computational Optimization and Applications*, **81**, 479-518. https://doi.org/10.1007/s10589-021-00338-8

[7] Nesterov, Y. (1983) A Method for Unconstrained Convex Minimization Problem with the Rate of Convergence $O(1/k^2)$. *Doklady Akademii Nauk*, **269**, 543.

[8] Beck, A. and Teboulle, M. (2009) A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, **2**, 183-202. https://doi.org/10.1137/080716542

[9] Hestenes, M.R. (1969) Multiplier and Gradient Methods. *Journal of Optimization Theory and Applications*, **4**, 303-320. https://doi.org/10.1007/bf00927673

[10] Powell, M.J.D. (1969) A Method for Nonlinear Constraints in Minimization Problems. In: Fletcher, R., Ed., *Optimization*, Academic Press, 283-298.

[11] Bertsekas, D.P. (2014) Constrained Optimization and Lagrange Multiplier Methods. Academic Press.

[12] Bertsekas, D.P. (1997) Nonlinear Programming. *Journal of the Operational Research Society*, **48**, 334-334. https://doi.org/10.1057/palgrave.jors.2600425

[13] Nocedal, J. and Wright, S.J. (1999) Numerical Optimization. Springer.

[14] He, B. and Yuan, X. (2010) On the Acceleration of Augmented Lagrangian Method for Linearly Constrained Optimization. Optimization Online.

[15] Ke, Y. and Ma, C. (2017) An Accelerated Augmented Lagrangian Method for Linearly Constrained Convex Programming with the Rate of Convergence $O(1/k^2)$. *Applied Mathematics—A Journal of Chinese Universities*, **32**, 117-126. https://doi.org/10.1007/s11766-017-3381-z