

On the acceleration of augmented Lagrangian method for linearly constrained optimization

Bingsheng He* and Xiaoming Yuan †

October 11, 2010

Abstract. The classical augmented Lagrangian method (ALM) plays a fundamental role in algorithmic development of constrained optimization. In this paper, we mainly show that Nesterov's influential acceleration techniques can be applied to accelerate ALM, thus yielding an accelerated ALM whose iteration-complexity is $O(1/k^2)$ for linearly constrained convex programming. As a by-product, we also show easily that the convergence rate of the original ALM is $O(1/k)$.

Keywords. Convex programming, augmented Lagrangian method, acceleration.

1 Introduction

The classical augmented Lagrangian method (ALM), or well-known as the method of multipliers, has been playing a fundamental role in the algorithmic development of constrained optimization ever since its presence in [2] and [9]. The existing literature about ALM is too many to be listed, and we only refer to [1, 8] for its comprehensive study.

In this paper, we restrict our discussion into the convex minimization with linear equation constrains:

$$(P) \quad \min \{f(x) \mid Ax = b, x \in \mathcal{X}\}, \quad (1.1)$$

where $f(x) : \mathfrak{R}^n \rightarrow \mathfrak{R}$ is a differentiable convex function, $A \in \mathfrak{R}^{m \times n}$ and $b \in \mathfrak{R}^m$ and \mathcal{X} is a convex closed set in \mathfrak{R}^n . Throughout we assume that the solution set of (1.1) denoted by \mathcal{X}^* is not empty.

Note that the Lagrange function of the problem (1.1) is

$$L(x, \lambda) = f(x) - \lambda^T(Ax - b), \quad (1.2)$$

where $\lambda \in \mathfrak{R}^m$ is the Lagrange multiplier. Then, the dual problem of (1.1) is

$$(D) \quad \begin{aligned} & \max_{x \in \mathcal{X}, \lambda \in \mathfrak{R}^m} L(x, \lambda) \\ & \text{s. t. } (x' - x)^T \nabla_x L(x, \lambda) \geq 0, \forall x' \in \mathcal{X}. \end{aligned} \quad (1.3)$$

We denote the solution set of (1.3) by $\mathcal{X}^* \times \Lambda^*$. As analyzed in [1], the ALM merges the penalty idea with the primal-dual and Lagrangian philosophy, and each of its iteration consists of the task of minimizing the augmented Lagrangian function of (1.1) and the task of updating the Lagrange multiplier. More specifically, starting with $\lambda^0 \in \mathfrak{R}^m$, the k -th iteration of ALM for (1.1) is

$$\begin{cases} x^{k+1} = \text{Argmin} \{f(x) - (\lambda^k)^T(Ax - b) + \frac{\beta}{2}\|Ax - b\|^2 \mid x \in \mathcal{X}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} - b), \end{cases} \quad (1.4)$$

*Department of Mathematics and National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, 210093, China. This author was supported by the NSFC Grant 10971095 and the NSF of Jiangsu Province Grant BK2008255. Email: hebma@nju.edu.cn

†Department of Mathematics, Hong Kong Baptist University, Hong Kong, China. This author was supported in part by HKRGC 203009. Email: xmyuan@hkbu.edu.hk

where $\beta > 0$ is the penalty parameter for the violation of the linear constraints. We refer to [10] for the relevance of ALM with the classical proximal point algorithm, which was originally proposed in [4] and concretely developed in [11]. Note that among significant differences of ALM from penalty methods is that the penalty parameter β can be fixed and it is not necessary to be forced to infinity, see e.g. [8]. In this paper, we use a symmetric positive definite matrix H_k to denote the penalty parameter, indicating the eligibility of adjusting values of this parameter dynamically even though the specific strategy of this adjustment will not be addressed. More specifically, let $\{H_k\}$ be a given series of $m \times m$ symmetric positive definite matrices and it satisfy

$$H_k - H_{k+1} \preceq 0, \quad \forall k > 0.$$

Then, the k -th iteration of ALM with matrix penalty parameter for (1.1) can be written as

$$\begin{cases} x^{k+1} = \text{Argmin} \{f(x) - (\lambda^k)^T(Ax - b) + \frac{1}{2}\|Ax - b\|_{H_k}^2 \mid x \in \mathcal{X}\}, \\ \lambda^{k+1} = \lambda^k - H_k(Ax^{k+1} - b). \end{cases} \quad (1.5)$$

Inspired by the attractive analysis of iteration-complexity for some gradient methods initiated mainly by Nesterov (see e.g. [5, 6, 7]), in this paper, we are interested in analyzing the iteration-complexity of the ALM and discussing the possibility of accelerating ALM with Nesterov's acceleration schemes. More specifically, in Section 2 we shall first show that the iteration-complexity of the ALM is $O(1/k)$ in terms of the objective residual of the associated Lagrange function of (1.1). Then, in Section 3, with the acceleration scheme in [6], we propose an accelerated ALM whose iteration-complexity is $O(1/k^2)$. Finally, some conclusions are given in Section 4.

2 The complexity of ALM

In this section, we mainly show that the iteration-complexity of the classical ALM is $O(1/k)$ in terms of the objective residual of the associated Lagrange function $L(x, \lambda)$ defined in (1.2). Before that, we need to justify the rationale of estimating the convergence rate of ALM in terms of the objective residual of $L(x, \lambda)$ and prove some properties of the sequence generated by ALM which are critical for complexity analysis to be addressed later.

According to (1.3), a pair $(x, \lambda) \in \mathcal{X} \times \Re^m$ is dual feasible if and only if

$$(x' - x)^T(\nabla f(x) - A^T \lambda) \geq 0, \quad \forall x' \in \mathcal{X}. \quad (2.1)$$

Note that the minimization task regarding x^{k+1} in the ALM scheme (1.5) is characterized by the following variational inequality:

$$(x - x^{k+1})^T \{\nabla f(x^{k+1}) - A^T \lambda^k + A^T H_k(Ax^{k+1} - b)\} \geq 0, \quad \forall x \in \mathcal{X}.$$

Therefore, substituting the λ^{k+1} -related equation in (1.5) into the above variational inequality, we have

$$(x - x^{k+1})^T \{\nabla f(x^{k+1}) - A^T \lambda^{k+1}\} \geq 0, \quad \forall x \in \mathcal{X}. \quad (2.2)$$

In other words, the pair (x^{k+1}, λ^{k+1}) generated by the k -th iteration of ALM is feasible to the dual problem (1.3). On the other hand, a solution $(x^*, \lambda^*) \in \mathcal{X}^* \times \Lambda^*$ of (1.3) is also feasible. We thus have that the sequence $\{L(x^*, \lambda^*) - L(x^{k+1}, \lambda^{k+1})\}$ is non-negative. This explains the rationale of estimating the convergence rate of ALM in terms of the objective residual of $L(x, \lambda)$.

Now, we present some properties of the sequence generated by the ALM in the following lemmas. Despite that their proofs are elementary, these lemmas are critical for deriving the main results of iteration-complexity later.

Lemma 2.1. *For given λ^k , let (x^{k+1}, λ^{k+1}) be generated by the ALM (1.5). Then, for any feasible solution (x, λ) of the dual problem (1.3), we have*

$$L(x^{k+1}, \lambda^{k+1}) - L(x, \lambda) \geq \|\lambda^k - \lambda^{k+1}\|_{H_k^{-1}}^2 + (\lambda - \lambda^k)^T H_k^{-1} (\lambda^k - \lambda^{k+1}). \quad (2.3)$$

Proof. First, using the convexity of f we obtain

$$\begin{aligned} & L(x^{k+1}, \lambda^{k+1}) - L(x, \lambda) \\ &= f(x^{k+1}) - f(x) + \lambda^T(Ax - b) - (\lambda^{k+1})^T(Ax^{k+1} - b) \\ &\geq (x^{k+1} - x)^T \nabla f(x) + \lambda^T(Ax - b) - (\lambda^{k+1})^T(Ax^{k+1} - b). \end{aligned} \quad (2.4)$$

Since (x, λ) is a feasible solution of the dual problem and $x^{k+1} \in \mathcal{X}$, set $x' = x^{k+1}$ in (2.1), we obtain

$$(x^{k+1} - x)^T \nabla f(x) \geq (x^{k+1} - x)^T A^T \lambda = \lambda^T A(x^{k+1} - x).$$

Substituting the last inequality in the right-hand side of (2.4), we obtain

$$\begin{aligned} L(x^{k+1}, \lambda^{k+1}) - L(x, \lambda) &\geq \lambda^T A(x^{k+1} - x) + \lambda^T(Ax - b) - (\lambda^{k+1})^T(Ax^{k+1} - b) \\ &= (\lambda - \lambda^{k+1})^T(Ax^{k+1} - b) \quad (\text{using (1.5)}) \\ &= (\lambda - \lambda^{k+1})^T H_k^{-1}(\lambda^k - \lambda^{k+1}) \\ &= \|\lambda^k - \lambda^{k+1}\|_{H_k^{-1}}^2 + (\lambda - \lambda^k)^T H_k^{-1}(\lambda^k - \lambda^{k+1}). \end{aligned}$$

The assertion of this lemma is proved. \square

Lemma 2.2. For given λ^k , let λ^{k+1} be generated by the ALM (1.5). Then we have

$$\begin{aligned} \|\lambda^{k+1} - \lambda^*\|_{H_k^{-1}}^2 &\leq \|\lambda^k - \lambda^*\|_{H_k^{-1}}^2 - \|\lambda^k - \lambda^{k+1}\|_{H_k^{-1}}^2 \\ &\quad - 2(L(x^*, \lambda^*) - L(x^{k+1}, \lambda^{k+1})), \quad \forall (x^*, \lambda^*) \in \mathcal{X}^* \times \Lambda^*. \end{aligned} \quad (2.5)$$

Proof. Since (x^*, λ^*) is dual feasible, by setting $(x, \lambda) = (x^*, \lambda^*)$ in (2.3), we obtain

$$(\lambda^k - \lambda^*)^T H_k^{-1}(\lambda^k - \lambda^{k+1}) \geq \|\lambda^k - \lambda^{k+1}\|_{H_k^{-1}}^2 + (L(x^*, \lambda^*) - L(x^{k+1}, \lambda^{k+1})).$$

Using the above inequality and by a manipulation, we obtain

$$\begin{aligned} \|\lambda^{k+1} - \lambda^*\|_{H_k^{-1}}^2 &= \|(\lambda^k - \lambda^*) - (\lambda^k - \lambda^{k+1})\|_{H_k^{-1}}^2 \\ &= \|\lambda^k - \lambda^*\|_{H_k^{-1}}^2 - 2(\lambda^k - \lambda^*)^T H_k^{-1}(\lambda^k - \lambda^{k+1}) + \|\lambda^k - \lambda^{k+1}\|_{H_k^{-1}}^2 \\ &\leq \|\lambda^k - \lambda^*\|_{H_k^{-1}}^2 - \|\lambda^k - \lambda^{k+1}\|_{H_k^{-1}}^2 - 2(L(x^*, \lambda^*) - L(x^{k+1}, \lambda^{k+1})). \end{aligned}$$

The assertion of this lemma is proved. \square

The following theorem implies the global convergence of ALM (1.5).

Theorem 2.3. Let (x^{k+1}, λ^{k+1}) be generated by the ALM (1.5). Then for any $k \geq 1$, we have

$$L(x^{k+1}, \lambda^{k+1}) \geq L(x^k, \lambda^k) + \|\lambda^k - \lambda^{k+1}\|_{H_k^{-1}}^2, \quad (2.6)$$

and

$$\|\lambda^{k+1} - \lambda^*\|_{H_{k+1}^{-1}}^2 \leq \|\lambda^k - \lambda^*\|_{H_k^{-1}}^2 - \|\lambda^k - \lambda^{k+1}\|_{H_k^{-1}}^2. \quad (2.7)$$

Moreover, if $H_k \equiv H$, we have

$$\|Ax^{k+1} - b\|_H^2 \leq \|Ax^k - b\|_H^2 - \|A(x^k - x^{k+1})\|_H^2. \quad (2.8)$$

Proof. The first assertion (2.6) of this theorem is derived immediately from (2.3). Since $L(x^{k+1}, \lambda^{k+1}) \leq L(x^*, \lambda^*)$, it follows from (2.5) that

$$\|\lambda^{k+1} - \lambda^*\|_{H_k^{-1}}^2 \leq \|\lambda^k - \lambda^*\|_{H_k^{-1}}^2 - \|\lambda^k - \lambda^{k+1}\|_{H_k^{-1}}^2.$$

Because $H_{k+1}^{-1} \preceq H_k^{-1}$, the second assertion (2.7) follows from the last inequality directly.

Now, we start to prove the third assertion (2.8). Setting $x = x^k$ in (2.2), we obtain

$$(x^k - x^{k+1})^T (\nabla f(x^{k+1}) - A^T \lambda^{k+1}) \geq 0.$$

Similarly, we have

$$(x^{k+1} - x^k)^T (\nabla f(x^k) - A^T \lambda^k) \geq 0.$$

Adding the above two inequalities and using the monotonicity of ∇f , we obtain

$$(x^k - x^{k+1})^T A^T (\lambda^k - \lambda^{k+1}) \geq 0.$$

By using $\lambda^{k+1} = \lambda^k - H_k(Ax^{k+1} - b)$ (and the assumption $H_k \equiv H$), the last inequality that

$$(x^k - x^{k+1})^T A^T H(Ax^{k+1} - b) \geq 0.$$

Using the above inequality in the identity

$$\|Ax^k - b\|_H^2 = \|Ax^{k+1} - b\|_H^2 + \|A(x^k - x^{k+1})\|_H^2 + 2(Ax^{k+1} - b)^T HA(x^k - x^{k+1}),$$

we obtain

$$\|Ax^k - b\|_H^2 \geq \|Ax^{k+1} - b\|_H^2 + \|A(x^k - x^{k+1})\|_H^2,$$

and thus the third assertion (2.8) is proved. \square

Remark 2.4. The inequality (2.7) essentially implies the global convergence of the ALM (1.5) with dynamically-adjusted matrix penalty parameter. In fact, it follows from (2.7) that

$$\sum_{l=0}^k \|\lambda^l - \lambda^{l+1}\|_{H_l^{-1}}^2 \leq \|\lambda^0 - \lambda^*\|_{H_0^{-1}}^2,$$

which instantly implies that

$$\lim_{k \rightarrow \infty} \|\lambda^k - \lambda^{k+1}\|_{H_k^{-1}}^2 = 0.$$

In the following we show that the sequence of function value $\{L(x^k, \lambda^k)\}$ converges to the optimal value $L(x^*, \lambda^*)$ at a rate of convergence that is no worse than $O(1/k)$. Hence, the iteration-complexity of the ALM (1.5) is shown to be $O(1/k)$ in terms of the objective residual of the Lagrange function $L(x, \lambda)$.

Theorem 2.5. *Let (x^k, λ^k) be generated by the ALM (1.5). Then, for any $k \geq 1$, we have*

$$L(x^*, \lambda^*) - L(x^k, \lambda^k) \leq \frac{\|\lambda^0 - \lambda^*\|_{H_0^{-1}}^2}{2k}, \quad \forall (x^*, \lambda^*) \in \mathcal{X}^* \times \Lambda^*. \quad (2.9)$$

Proof. Due to $H_{k+1} \succeq H_k$, it follows from Lemma 2.2 that, for all $j \geq 0$, we have

$$\begin{aligned} & 2(L(x^{j+1}, \lambda^{j+1}) - L(x^*, \lambda^*)) \\ & \geq \|\lambda^{j+1} - \lambda^*\|_{H_{j+1}^{-1}}^2 - \|\lambda^j - \lambda^*\|_{H_j^{-1}}^2 + \|\lambda^j - \lambda^{j+1}\|_{H_j^{-1}}^2, \quad \forall (x^*, \lambda^*) \in \mathcal{X}^* \times \Lambda^*. \end{aligned}$$

Using the fact that $L(x^{j+1}, \lambda^{j+1}) - L(x^*, \lambda^*) \leq 0$ and summing the above inequality over $j = 0, \dots, k-1$, we obtain

$$\begin{aligned} & 2 \left(\sum_{j=0}^{k-1} L(x^{j+1}, \lambda^{j+1}) - kL(x^*, \lambda^*) \right) \\ & \geq \|\lambda^k - \lambda^*\|_{H_k^{-1}}^2 - \|\lambda^0 - \lambda^*\|_{H_0^{-1}}^2 + \sum_{j=1}^{k-1} \|\lambda^j - \lambda^{j+1}\|_{H_j^{-1}}^2. \end{aligned} \quad (2.10)$$

By using Lemma 2.1 for $k = j$ ($j \geq 1$) and $(x, \lambda) = (x^j, \lambda^j)$, we get

$$L(x^{j+1}, \lambda^{j+1}) - L(x^j, \lambda^j) \geq \|\lambda^j - \lambda^{j+1}\|_{H_j^{-1}}^2.$$

Multiplying the last inequality by $2j$ and summing over $j = 0, \dots, k-1$, it follows that

$$2 \sum_{j=1}^{k-1} \left((j+1)L(x^{j+1}, \lambda^{j+1}) - jL(x^j, \lambda^j) - L(x^{j+1}, \lambda^{j+1}) \right) \geq \sum_{j=1}^{k-1} 2j \|\lambda^j - \lambda^{j+1}\|_{H_j^{-1}}^2,$$

which can be simplified into

$$2 \left(kL(x^k, \lambda^k) - \sum_{j=0}^{k-1} L(x^{j+1}, \lambda^{j+1}) \right) \geq \sum_{j=0}^{k-1} 2j \|\lambda^j - \lambda^{j+1}\|_{H_j^{-1}}^2. \quad (2.11)$$

Adding (2.10) and (2.11), we get

$$2k(L(x^k, \lambda^k) - L(x^*, \lambda^*)) \geq \|\lambda^k - \lambda^*\|_{H_k^{-1}}^2 - \|\lambda^0 - \lambda^*\|_{H_0^{-1}}^2 + \sum_{j=0}^{k-1} (2j+1) \|\lambda^j - \lambda^{j+1}\|_{H_j^{-1}}^2,$$

and hence it follows that

$$L(x^*, \lambda^*) - L(x^k, \lambda^k) \leq \frac{\|\lambda^0 - \lambda^*\|_{H_0^{-1}}^2}{2k}.$$

The proof is complete. \square

3 An accelerated ALM

In this section, we show that the classical ALM (1.5) can be accelerated by some influential acceleration techniques initialized by Nesterov in [6]. As a result, an accelerated ALM with the convergence rate $O(1/k^2)$ for solving (1.3) is proposed.

For the convenience of presenting the accelerated ALM, from now on we use $(\tilde{x}^k, \tilde{\lambda}^k)$, rather than (x^{k+1}, λ^{k+1}) , to denote the iterate generated by the ALM scheme (1.5). Namely, with the given λ^k , the new iterate generated by ALM for (1.1) is $(\tilde{x}^k, \tilde{\lambda}^k)$:

$$\begin{cases} \tilde{x}^k = \text{Argmin} \{ f(x) - (\lambda^k)^T (Ax - b) + \frac{1}{2} \|Ax - b\|_{H_k}^2 \mid x \in \mathcal{X} \}, \\ \tilde{\lambda}^k = \lambda^k - H_k (A\tilde{x}^k - b). \end{cases} \quad (3.1)$$

Accordingly, Lemmas 2.1 and 2.2 can be rewritten into the following lemmas.

Lemma 3.1. *For given λ^k , let $(\tilde{x}^k, \tilde{\lambda}^k)$ be generated by the ALM (3.1). Then, for any feasible solution (x, λ) of the dual problem (1.3), we have*

$$L(\tilde{x}^k, \tilde{\lambda}^k) - L(x, \lambda) \geq \|\lambda^k - \tilde{\lambda}^k\|_{H_k^{-1}}^2 + (\lambda - \lambda^k)^T H_k^{-1} (\lambda^k - \tilde{\lambda}^k). \quad (3.2)$$

Lemma 3.2. *For given λ^k , let $(\tilde{x}^k, \tilde{\lambda}^k)$ be generated by the ALM (3.1). Then we have*

$$\begin{aligned} \|\tilde{\lambda}^k - \lambda^*\|_{H_k^{-1}}^2 &\leq \|\lambda^k - \lambda^*\|_{H_k^{-1}}^2 - \|\lambda^k - \tilde{\lambda}^k\|_{H_k^{-1}}^2 \\ &\quad - 2(L(x^*, \lambda^*) - L(\tilde{x}^k, \tilde{\lambda}^k)), \quad \forall (x^*, \lambda^*) \in \mathcal{X}^* \times \Lambda^*. \end{aligned} \quad (3.3)$$

Then, the accelerated ALM for (1.1) is as follows.

An accelerated augmented Lagrangian method (AALM)
Step 0. Take $\tilde{\lambda}^0 \in \mathfrak{R}^m$. Set $\lambda^1 = \tilde{\lambda}^0$ and $t_1 = 1$.
Step k. Let $(\tilde{x}^k, \tilde{\lambda}^k)$ be generated by the original ALM (3.1).
Set

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}, \quad (3.4a)$$

and

$$\lambda^{k+1} = \tilde{\lambda}^k + \left(\frac{t_k - 1}{t_{k+1}} \right) (\tilde{\lambda}^k - \tilde{\lambda}^{k-1}) + \left(\frac{t_k}{t_{k+1}} \right) (\tilde{\lambda}^k - \lambda^k). \quad (3.4b)$$

We first propose some lemmas before the main result.

Lemma 3.3. *The sequence $\{t_k\}$ generated by (3.4a) with $t_1 = 1$ satisfies*

$$t_k \geq (k+1)/2, \quad \forall k \geq 1. \quad (3.5)$$

Proof. Elementary by induction. \square

For the coming analysis, we use the notations

$$v_k := L(x^*, \lambda^*) - L(\tilde{x}^k, \tilde{\lambda}^k) \quad \text{and} \quad u^k := t_k(2\tilde{\lambda}^k - \lambda^k - \tilde{\lambda}^{k-1}) + \tilde{\lambda}^{k-1} - \lambda^*. \quad (3.6)$$

Lemma 3.4. *The sequences $\{\lambda^k\}$ and $\{\tilde{\lambda}^k\}$ generated by the proposed AALM satisfy*

$$4t_k^2 v_k - 4t_{k+1}^2 v_{k+1} \geq \|u^{k+1}\|_{H_{k+1}^{-1}}^2 - \|u^k\|_{H_{k+1}^{-1}}^2, \quad \forall k \geq 1, \quad (3.7)$$

where v_k and u^k are defined in (3.6).

Proof. By using Lemma 3.1 for $k+1$, setting $(x, \lambda) = (\tilde{x}^k, \lambda^k)$ and $(x, \lambda) = (x^*, \lambda^*)$, we get

$$L(\tilde{x}^{k+1}, \tilde{\lambda}^{k+1}) - L(\tilde{x}^k, \tilde{\lambda}^k) \geq \|\lambda^{k+1} - \tilde{\lambda}^{k+1}\|_{H_{k+1}^{-1}}^2 + (\tilde{\lambda}^k - \lambda^{k+1})^T H_{k+1}^{-1} (\lambda^{k+1} - \tilde{\lambda}^{k+1}),$$

and

$$L(\tilde{x}^{k+1}, \tilde{\lambda}^{k+1}) - L(x^*, \lambda^*) \geq \|\lambda^{k+1} - \tilde{\lambda}^{k+1}\|_{H_{k+1}^{-1}}^2 + (\lambda^* - \lambda^{k+1})^T H_{k+1}^{-1} (\lambda^{k+1} - \tilde{\lambda}^{k+1}),$$

respectively. Using the definition of v_k , the last two inequalities can be written as

$$v_k - v_{k+1} \geq \|\lambda^{k+1} - \tilde{\lambda}^{k+1}\|_{H_{k+1}^{-1}}^2 + (\tilde{\lambda}^k - \lambda^{k+1})^T H_{k+1}^{-1} (\lambda^{k+1} - \tilde{\lambda}^{k+1}), \quad (3.8)$$

and

$$-v_{k+1} \geq \|\lambda^{k+1} - \tilde{\lambda}^{k+1}\|_{H_{k+1}^{-1}}^2 + (\lambda^* - \lambda^{k+1})^T H_{k+1}^{-1} (\lambda^{k+1} - \tilde{\lambda}^{k+1}). \quad (3.9)$$

To get a relation between v_k and v_{k+1} , we multiply (3.8) by $(t_{k+1} - 1)$ and add it to (3.9):

$$\begin{aligned} & (t_{k+1} - 1)v_k - t_{k+1}v_{k+1} \\ & \geq t_{k+1}\|\tilde{\lambda}^{k+1} - \lambda^{k+1}\|_{H_{k+1}^{-1}}^2 + (\tilde{\lambda}^{k+1} - \lambda^{k+1})^T H_{k+1}^{-1} (t_{k+1}\lambda^{k+1} - (t_{k+1} - 1)\tilde{\lambda}^k - \lambda^*). \end{aligned}$$

Multiplying the last inequality by t_{k+1} and using

$$t_k^2 = t_{k+1}^2 - t_{k+1} \quad (\text{and thus } t_{k+1} = (1 + \sqrt{1 + 4t_k^2})/2 \text{ as in (3.4a)}),$$

which yields

$$\begin{aligned} & t_k^2 v_k - t_{k+1}^2 v_{k+1} \\ & \geq \|t_{k+1}(\tilde{\lambda}^{k+1} - \lambda^{k+1})\|_{H_{k+1}^{-1}}^2 + t_{k+1}(\tilde{\lambda}^{k+1} - \lambda^{k+1})^T H_{k+1}^{-1} (t_{k+1}\lambda^{k+1} - (t_{k+1} - 1)\tilde{\lambda}^k - \lambda^*) \\ & = (t_{k+1}(\tilde{\lambda}^{k+1} - \lambda^{k+1}))^T H_{k+1}^{-1} (t_{k+1}\tilde{\lambda}^{k+1} - (t_{k+1} - 1)\tilde{\lambda}^k - \lambda^*). \end{aligned} \quad (3.10)$$

Use the identity

$$(b-a)^T H_{k+1}^{-1} (b-c) = \frac{1}{4} \|(2b-a) - c\|_{H_{k+1}^{-1}}^2 - \frac{1}{4} \|a-c\|_{H_{k+1}^{-1}}^2$$

(since $x^T y = \frac{1}{4} \|x+y\|^2 - \frac{1}{4} \|x-y\|^2$) to the right-hand side of (3.10) with

$$a := t_{k+1}\lambda^{k+1}, \quad b := t_{k+1}\tilde{\lambda}^{k+1}, \quad c := (t_{k+1} - 1)\tilde{\lambda}^k + \lambda^*,$$

we get

$$\begin{aligned} t_k^2 v_k - t_{k+1}^2 v_{k+1} & \geq \frac{1}{4} \|t_{k+1}(2\tilde{\lambda}^{k+1} - \lambda^{k+1} - \tilde{\lambda}^k) + \tilde{\lambda}^k - \lambda^*\|_{H_{k+1}^{-1}}^2 \\ & \quad - \frac{1}{4} \|t_{k+1}(\lambda^{k+1} - \tilde{\lambda}^k) + \tilde{\lambda}^k - \lambda^*\|_{H_{k+1}^{-1}}^2. \end{aligned}$$

Using the notation of $u^k := t_k(2\tilde{\lambda}^k - \lambda^k - \tilde{\lambda}^{k-1}) + \tilde{\lambda}^{k-1} - \lambda^*$ (see (3.6)), the last inequality can be written as

$$4t_k^2 v_k - 4t_{k+1}^2 v_{k+1} \geq \|u^{k+1}\|_{H_{k+1}^{-1}}^2 - \|t_{k+1}(\lambda^{k+1} - \tilde{\lambda}^k) + \tilde{\lambda}^k - \lambda^*\|_{H_{k+1}^{-1}}^2. \quad (3.11)$$

In order to write the inequality (3.11) in the form (3.7), we need only to set

$$t_{k+1}(\lambda^{k+1} - \tilde{\lambda}^k) + \tilde{\lambda}^k - \lambda^* = t_k(2\tilde{\lambda}^k - \lambda^k - \tilde{\lambda}^{k-1}) + \tilde{\lambda}^{k-1} - \lambda^*.$$

From the last equality we obtain

$$\lambda^{k+1} = \tilde{\lambda}^k + \left(\frac{t_k - 1}{t_{k+1}}\right)(\tilde{\lambda}^k - \tilde{\lambda}^{k-1}) + \left(\frac{t_k}{t_{k+1}}\right)(\tilde{\lambda}^k - \lambda^k).$$

This is just the form (3.4b) in the accelerated multi-step version of the ALM and the lemma is proved. \square

Corollary 3.5. *Let v_k and u^k be defined in (3.6). Then, we have*

$$4t_k^2 v_k \leq 4t_1^2 v_1 + \|u^1\|_{H_1^{-1}}^2, \quad \forall k \geq 1. \quad (3.12)$$

Proof. Again, because $H_{k+1}^{-1} \preceq H_k^{-1}$, from (3.7) we obtain

$$4t_k^2 v_k - 4t_{k+1}^2 v_{k+1} \geq \|u^{k+1}\|_{H_{k+1}^{-1}}^2 - \|u^k\|_{H_k^{-1}}^2.$$

Since $\{v_k\}$ is a non-negative sequence, the last inequality implies (3.12) immediately. \square

Now, we are ready to show the fact that the iteration-complexity of the proposed AALM is $O(1/k^2)$.

Theorem 3.6. *Let $\{\tilde{\lambda}^k\}$ and $\{\lambda^k\}$ be generated by the proposed AALM. Then, for any $k \geq 1$, we have*

$$L(x^*, \lambda^*) - L(\tilde{x}^k, \tilde{\lambda}^k) \leq \frac{\|\lambda^1 - \lambda^*\|_{H_1^{-1}}^2}{(k+1)^2}, \quad \forall (x^*, \lambda^*) \in \mathcal{X}^* \times \Lambda^*. \quad (3.13)$$

Proof. Using the definition of v_k in (3.6), it follows from (3.12) that

$$L(x^*, \lambda^*) - L(\tilde{x}^k, \tilde{\lambda}^k) = v_k \leq \frac{4t_1^2 v_1 + \|u^1\|_{H_1^{-1}}^2}{4t_k^2}.$$

Combining with the fact $t_k \geq (k+1)/2$ (see (3.5)), it yields

$$L(x^*, \lambda^*) - L(\tilde{x}^k, \tilde{\lambda}^k) \leq \frac{4t_1^2 v_1 + \|u^1\|_{H_1^{-1}}^2}{(k+1)^2}. \quad (3.14)$$

Since $t_1 = 1$, and using the definition of u^k given in (3.6), we have

$$4t_1^2 v_1 = 4v_1 = 4(L(x^*, \lambda^*) - L(\tilde{x}^1, \tilde{\lambda}^1)), \quad \|u^1\|_{H_1^{-1}}^2 = \|2\tilde{\lambda}^1 - \lambda^1 - \lambda^*\|_{H_1^{-1}}^2. \quad (3.15)$$

By using (3.3), we have

$$4(L(x^*, \lambda^*) - L(\tilde{x}^1, \tilde{\lambda}^1)) \leq 2\|\lambda^1 - \lambda^*\|_{H_1^{-1}}^2 - 2\|\tilde{\lambda}^1 - \lambda^*\|_{H_1^{-1}}^2 - 2\|\tilde{\lambda}^1 - \lambda^1\|_{H_1^{-1}}^2. \quad (3.16)$$

Use the identity

$$2\|a - c\|^2 - 2\|b - c\|^2 - 2\|b - a\|^2 = \|a - c\|^2 - \|(b - a) + (b - c)\|^2$$

to the right-hand side of (3.16) with

$$a := \lambda^1, \quad b := \tilde{\lambda}^1, \quad c := \lambda^*,$$

we get

$$4(L(x^*, \lambda^*) - L(\tilde{x}^1, \tilde{\lambda}^1)) \leq \|\lambda^1 - \lambda^*\|_{H_1^{-1}}^2 - \|2\tilde{\lambda}^1 - \lambda^1 - \lambda^*\|_{H_1^{-1}}^2. \quad (3.17)$$

Consequently, it follows from (3.15) and (3.17) that

$$4t_1^2 v_1 + \|u^1\|_{H_1^{-1}}^2 \leq \|\lambda^1 - \lambda^*\|_{H_1^{-1}}^2.$$

Substituting it in (3.14), the assertion is proved. \square

According to Theorem 3.6, for obtaining an ε -optimal solution of (1.3) (denoted by $(\tilde{x}, \tilde{\lambda})$) in the sense that $L(x^*, \lambda^*) - L(\tilde{x}, \tilde{\lambda}) \leq \varepsilon$, the number of iterations required by the proposed accelerated ALM is at most $\lceil C/\sqrt{\varepsilon} - 1 \rceil$ where $C = \|\lambda^1 - \lambda^*\|_{H_1^{-1}}^2$.

4 Conclusions

In this paper, we first show that the iteration-complexity of the classical augmented Lagrangian method (ALM) is $O(1/k)$ for solving linearly constrained convex programming. Then, we show that the ALM can be accelerated by applying Nesterov's acceleration techniques, and the iteration-complexity of the yielded accelerated ALM is $O(1/k^2)$. In the future, we will investigate (a) the complexity of inexact ALM where the subproblems are solved approximately subject to certain criteria, as [3]; (b) the complexity of some ALM-based methods, e.g. the well-known alternating direction method for solving separable convex programming with linear constraints.

References

- [1] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Method*, Academic Press, New York, 1982.
- [2] M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appli., 4 (1969), pp. 303-320.
- [3] G. H. Lan and R. D. C. Monteiro, *Iteration-complexity of first-order augmented Lagrangian methods for convex programming*, manuscript, 2009.
- [4] B. Martinet, Regularisation, d'inéquations variationnelles par approximations succesives, *Rev. Francaise d'Inform. Recherche Oper.*, 4 (1970), pp. 154-159.
- [5] A. S. Nemirovsky and D. B. Yudin, *Problem Complexity and Method Efficiency in Optimization*, Wiley-Interscience Series in Discrete Mathematics, John Wiley & Sons, New York, 1983.
- [6] Y. E. Nesterov, *A method for solving the convex programming problem with convergence rate $O(1/k^2)$* , Dokl. Akad. Nauk SSSR, 269 (1983), pp. 543-547.
- [7] Y. E. Nesterov, *Gradient methods for minimizing composite objective function*, CORE report 2007; available at <http://www.ecore.be/DPs/dp-1191313936.pdf>.
- [8] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer Verlag, 1999.
- [9] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, In Optimization edited by R. Fletcher, pp. 283-298, Academic Press, New York, 1969.
- [10] R.T. Rockafellar, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Math. Oper. Res. 1 (1976), pp. 97-116.
- [11] R.T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM, J. Control Optim. 14 (1976), pp. 877-898.