AN AUGMENTED LAGRANGIAN METHOD FOR CONIC CONVEX PROGRAMMING

N. S. AYBAT* AND G. IYENGAR†

Abstract. We propose a new first-order augmented Lagrangian algorithm ALCC for solving convex conic programs of the form

$$\min\{\rho(x)+\gamma(x) : Ax-b\in\mathcal{K},\ x\in\chi\},$$

where $\rho:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$, $\gamma:\mathbb{R}^n\to\mathbb{R}$ are closed, convex functions, and γ has a Lipschitz continuous gradient, $A\in\mathbb{R}^{m\times n}$, $\mathcal{K}\subset\mathbb{R}^m$ is a closed convex cone, and $\chi\subset\mathbf{dom}(\rho)$ is a "simple" convex compact set such that optimization problems of the form $\min\{\rho(x)+\|x-\bar{x}\|_2^2 : x\in\chi\}$ can be efficiently solved. We show that any limit point of the primal ALCC iterates is an optimal solution of the conic convex problem, and the dual ALCC iterates have a unique limit point that is a Karush-Kuhn-Tucker (KKT) point of the conic program. We also show that for *any* $\epsilon>0$, the primal ALCC iterates are ϵ-feasible and ϵ-optimal after $\mathcal{O}(\log(\epsilon^{-1}))$ iterations which require solving $\mathcal{O}(\epsilon^{-1}\log(\epsilon^{-1}))$ problems of the form $\min_x\{\rho(x)+\|x-\bar{x}\|_2^2 :\ x\in\chi\}$.

1. Introduction.

In this paper we propose an inexact augmented Lagrangian algorithm (ALCC) for solving conic convex problems of the form

$$(P) : \min\left\{\rho(x)+\gamma(x) : Ax-b\in\mathcal{K},\ x\in\chi\right\}, \tag{1.1}$$

where $\rho:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$, $\gamma:\mathbb{R}^n\to\mathbb{R}$ are proper, closed, convex functions, and γ has a Lipschitz continuous gradient $\nabla\gamma$ with the Lipschitz constant L_γ, $A\in\mathbb{R}^{m\times n}$, $\mathcal{K}\subset\mathbb{R}^m$ is a nonempty, closed, convex cone, and $\chi\subset\mathbf{dom}(\rho)$ is a "simple" compact set in the sense that the optimization problems of the form

$$\min_{x\in\chi}\left\{\rho(x)+\|x-\bar{x}\|_2^2\right\} \tag{1.2}$$

can be efficiently solved for any $\bar{x}\in\mathbb{R}^n$. Note that we do not require $A\in\mathbb{R}^{m\times n}$ to satisfy any additional regularity properties. For notational convenience, we set

$$p(x) := \rho(x)+\gamma(x).$$

In some problems, the compact set χ is explicitly present. For example, in a zero-sum game the decision x represents a mixed strategy and the set χ is a simplex. In others, χ may not be explicitly present, but one can formulate an equivalent problem where the vector of decision variables can be constrained to lie in a bounded feasible set without any loss of generality. For example, if γ is strongly convex, or if ρ is a norm and $\gamma(\cdot)\geq 0$, then the decision vector x can be restricted to lie in a appropriately defined norm ball centered at any feasible solution.

We assume that the following constraint qualification holds for (P).

ASSUMPTION 1.1. *The problem* (P) *in* (1.1) *has a Karush-Kuhn-Tucker (KKT) point, i.e., there exists* $y^*\in\mathcal{K}^*$ *such that* $g_0(y^*) := \inf\{p(x)-\langle y^*,\ Ax-b\rangle :\ x\in\chi\} = p^* > -\infty$*, where* p^* *denotes the optimal value of* (P) *and* $\mathcal{K}^*$ *denotes the dual cone corresponding to* $\mathcal{K}$*, i.e.,* $\mathcal{K}^* := \{y\in\mathbb{R}^m :\ \langle y,x\rangle\geq 0\ \forall x\in\mathcal{K}\}$.
Assumption 1.1 clearly holds whenever there exists $\tilde{x}\in\mathbf{relint}(\chi)$ such that $A\tilde{x}-b\in\mathbf{int}(\mathcal{K})$ [4].

1.1. Special cases.

Many important optimization problems are special cases of (1.1). Below, we briefly discuss some examples.

Min-max games with convex loss function: This problem is a generalization of the matrix game discussed in [11]. The decision maker can choose from n possible actions. Let $x\in\mathbb{R}^n_+$ denote a mixed strategy over the set of actions, i.e., $x\in\chi := \{x : \sum_{j=1}^n x_j = 1, x\geq \mathbf{0}\}$. Suppose the mixed strategy x must satisfy constraints of the form $Ax-b\in\mathcal{K}$. These constraints could be modeling average cost constraints. For example, one may have constraints of the form $Ax\leq b$, where $A\in\mathbb{R}^{m\times n}$ and A_{ij} denotes amount of resource i consumed by action j. One may also have constraints that restrict the total probability weight of some given subsets of actions.

*IE Department, Pennsylvania State University. Email: nsa10@psu.edu
†IEOR Department, Columbia University. Email: gi10@columbia.edu

The adversary has p possible actions. The expected loss to decision maker when she chooses the mixed strategy $x \in \mathbb{R}^n$ and the adversary chooses the mixed strategy $y \in \mathbb{R}^p$ is given by

$$\rho(x) + y^T Cx - \phi(y),$$

where ρ is a convex function, and ϕ is a strongly convex function. Then the decision maker's optimization problem that minimizes the expected worst case loss is given by

$$\min \left\{ \rho(x) + \gamma(x) : \ Ax - b \in \mathcal{K}, \ x \in \chi \right\}, \tag{1.3}$$

where

$$\gamma(x) = \max \left\{ y^T Cx - \phi(y) : \ \sum_{k=1}^{p} y_k = 1, \ y \geq \mathbf{0} \right\}. \tag{1.4}$$

From Danskin's theorem, it follows that $\nabla\gamma(x) = C^T y(x)$, where $y(x)$ denotes the unique minimizer in (1.4) for a given x. In [11], Nesterov showed that $\nabla\gamma$ is Lipschitz continuous with Lipschitz constant $\sigma_{\max}(C)^2/\tau$, where τ denotes the convexity parameter for the strongly convex function ϕ. Thus, it follows that the minimax optimization problem (1.3) is a special case of (1.1).

Problems with semidefinite constraints: Let $\mathcal{S}^m$ denote the set of $m \times m$ symmetric matrices, and let $\mathcal{S}^m_+$ denote the closed convex cone of $m \times m$ symmetric positive semidefinite matrices. A convex optimization problem with a linear matrix inequality constraint is of the form

$$\min \left\{ \rho(x) : \sum_{j=1}^{n} A_j x_j + B \in \mathcal{S}^m_+ \right\}, \tag{1.5}$$

where ρ is a convex function, $B \in \mathcal{S}^m$, and $A_j \in \mathcal{S}^m$ for $j = 1, \ldots, n$. Convex problems of the form (1.5) can model many applications in engineering, statistics and combinatorial optimization [4]. In most of these applications, either the constraints imply that the decision vector x is bounded, or one can often establish that the optimal solution lies in a norm-ball. In such cases, (1.5) is a special case of (1.1). Consider the ℓ_1-minimization problem of the form

$$\min \left\{ \|x\|_1 : \sum_{j=1}^{n} A_j x_j + B \in \mathcal{S}^m_+ \right\}. \tag{1.6}$$

Suppose a feasible solution x_0 for this problem is known. Then (1.6) is a special case of (1.1) with $\rho(x) = \|x\|_1$, $\gamma(\cdot) = 0$, $\mathcal{K} = \mathcal{S}^m_+$ and $\chi = \{x \in \mathbb{R}^n : \|x\|_1 \leq \|x_0\|_1\}$. The main bottleneck step in solving this problem using the ALCC algorithm reduces to the "shrinkage" problem of the form $\min\{\lambda\|x\|_1 + \|x - \bar{x}\|_2^2 : \ \|x\|_1 \leq \|x_0\|_1\}$ that can be solved very efficiently for any given $\bar{x} \in \mathbb{R}^n$ and $\lambda > 0$.

1.2. Notation. Let $S \subset \mathbb{R}^m$ be a nonempty, closed, convex set. Let $d_S : \mathbb{R}^m \to \mathbb{R}_+$ denote the function

$$d_S(\bar{x}) := \min_{x \in S} \|x - \bar{x}\|_2, \tag{1.7}$$

i.e., $d_S(\bar{x})$ denotes the ℓ_2-distance of the vector $\bar{x} \in \mathbb{R}^m$ to the set S. Let

$$\Pi_S(\bar{x}) := \operatorname{argmin}\{\|x - \bar{x}\|_2 : \ x \in S\}, \tag{1.8}$$

denote the ℓ_2-projection of the vector $\bar{x} \in \mathbb{R}^m$ onto the set S. Since $S \subset \mathbb{R}^m$ is a nonempty, closed, convex set, $\Pi_S(\cdot)$ is well defined. Moreover, $d_S(\bar{x}) = \|\bar{x} - \Pi_S(\bar{x})\|_2$.

1.3. New results. The main results of this paper are as follows:

(a) Every limit point of the sequence of ALCC primal iterates $\{x_k\}$ is an optimal solution of (1.1).
(b) The sequence of ALCC dual iterates $\{y_k\}$ converges to a KKT point of (1.1).

(c) For *all* $\epsilon > 0$, the primal ALCC iterates x_k are ϵ-feasible, i.e., $x_k \in \chi$ and $d_{\mathcal{K}}(Ax_k - b) \leq \epsilon$, and ϵ-optimal, i.e., $|p(x_k) - p^*| \leq \epsilon$ after at most $\mathcal{O}\left(\log\left(\epsilon^{-1}\right)\right)$ ALCC iterations that require solving at most $\mathcal{O}(\epsilon^{-1}\log(\epsilon^{-1}))$ problems of the form (1.2).

Since (1.1) is a conic convex programming problem, many special cases of (1.1) can be solved in polynomial time, at least in theory, using interior point methods. However, in practice, the interior point methods are not able to solve very large instances of (1.1) because the computational complexity of a matrix factorization step, which is essential in these methods, becomes prohibitive. On the other hand, the computational bottleneck in the ALCC algorithm is the projection (1.2). In many optimization problems that arise in applications, this projection can be solved very efficiently as is the case with noisy compressed sensing and matrix completion problems discussed in [2], and the convex optimization problems with semidefinite constraints discussed above. The convergence results above imply that the ALCC algorithm can solve very large instances of (1.1) very efficiently provided the corresponding projection (1.2) can be solved efficiently. The numerical results reported in [1, 2] for a special case of ALCC algorithm provide evidence that our proposed algorithm can be scaled to solve very large instances of the conic problem (1.1).

1.4. Previous work. Rockafellar [13] proposed an inexact augmented Lagrangian method to solve problems of the form

$$p^* = \min\left\{p(x) : f(x) \geq 0,\ x \in \chi\right\}, \tag{1.9}$$

where $\chi \subset \mathbb{R}^n$ is a closed convex set, $p : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a convex function and $f : \mathbb{R}^n \to \mathbb{R}^m$ such that each component $f_i(x)$ of $f = (f_1, \ldots, f_m)$ is a concave function for $i = 1, \ldots, m$. Rockafellar [13] defined the "penalty" Lagrangian

$$\tilde{\mathcal{L}}_\mu(x, y) := p(x) + \frac{\mu}{2}\left\|\left(\frac{y}{\mu} - f(x)\right)_+\right\|_2^2 - \frac{\|y\|_2^2}{2\mu}, \tag{1.10}$$

where $(\cdot)_+ := \max\{\cdot, \mathbf{0}\}$ and $\max\{\cdot,\cdot\}$ are componentwise operators, and μ is a fixed penalty parameter. Rockafellar [13] established that given $y_0 \in \mathbb{R}^m$, the primal-dual iterates sequences $\{x_k, y_k\} \subset \chi \times \mathbb{R}^m$ computed according to

$$\tilde{\mathcal{L}}_\mu(x_k, y_k) \leq \inf_{x\in\chi} \tilde{\mathcal{L}}_\mu(x, y_k) + \alpha_k, \tag{1.11}$$

$$y_{k+1} = \left(y_k + \mu f(x_k)\right)_+, \tag{1.12}$$

satisfy $\lim_{k\in\mathbb{Z}_+} p(x_k) = \bar{p}$ and $\limsup_{k\in\mathbb{Z}_+} f(x_k) \leq \mathbf{0}$ when (1.9) has a KKT point and the parameter sequence $\{\alpha_k\}$ satisfies the summability condition $\sum_{k=1}^{\infty}\sqrt{\mu\ \alpha_k} < \infty$. Martinet [9] later showed that the summability condition on parameter sequence $\{\alpha_k\}$ is not necessary. However, in both [9, 13] no iteration complexity result was given for the algorithm (1.11)–(1.12) when p was not continuously twice differentiable.

In this paper we show convergence rate results for an augmented Lagrangian algorithm where we allow penalty parameter μ to be a non-decreasing positive sequence $\{\mu_k\}$. After we had independently established these results, which are extensions of our previous results in [2], we became aware of a previous work by Rockafellar [14] where he proposed several different variants of the algorithm in (1.11)–(1.12) where μ could be updated between iterations. Rockafellar [14] established that for all non-decreasing positive multiplier sequences $\{\mu_k\}$ satisfying the summability condition $\sum_{k=1}^{\infty}\sqrt{\mu_k\ \alpha_k} < \infty$, $\{y_k\}$ is bounded and any limit point of $\{x_k\}$ is optimal to (1.9); moreover,

$$\max_{i=1,\ldots,m}\{f_i(x_k)\} \leq \frac{\|y_{k+1} - y_k\|_2}{\mu_k}, \quad p(x_k) - p^* \leq \frac{1}{2\mu_k}(\alpha_k + \|y_k\|_2^2). \tag{1.13}$$

Note that the results in [14] only provide an *upper bound* on the sub-optimality; no lower bound is provided. Since the iterates $\{x_k\}$ are only feasible in the limit, it is possible that $p(x_k) \ll p^*$ and establishing a lower bound on the sub-optimality is critical. Moreover, Rockafellar [14] does not discuss how to compute iterates satisfying (1.11) and assumes that a black-box oracle produces such iterates; consequently, there are no basic operation level complexity bounds in [14].

In this paper, we extend (1.9) to a conic convex program where $f(x) = Ax - b$, and $\mathcal{K}$ is a closed, convex cone. We show that primal ALCC iterates $\{x_k\} \subset \chi$ satisfies $d_{\mathcal{K}}(Ax_k - b) \leq \mathcal{O}(\mu_k^{-1})$ and $|p(x_k) - p^*| \leq \mathcal{O}(\mu_k^{-1})$, i.e. we provide *both* an upper and a lower bound, using an inexact stopping condition that is an extension of (1.11). ALCC algorithm calls an optimal first order method, such as FISTA [3], to compute an iterate x_k satisfying a stopping condition similar to (1.11). By carefully selecting the sub-optimality parameter sequence $\{\alpha_k\}$ and the penalty parameter sequence $\{\mu_k\}$, we are able to establish a bound on the number of generalized projections of the form (1.2) required to obtain an ϵ-feasible and ϵ-optimal solution to (1.1), and also provide an operation level complexity bound.

In [14], Rockafellar also provides an iteration complexity result for a different inexact augmented Lagrangian method. Given a non-increasing sequence $\{\alpha_k\}$ and a non-decreasing sequence $\{\mu_k\}$ such that $\sum_{k=1}^{\infty} \sqrt{\mu_k \ \alpha_k} < \infty$, the infeasiblity and suboptimality can be *upper* bounded (see (1.13)) when the duals $\{y_k\}$ are updated according to (1.12) and the primal iterates $\{x_k\}$ satisfy

$$\inf\{\|s\|_2 : \ s \in \partial\phi_k(x_k)\} \leq \sqrt{\frac{\alpha_k}{\mu_k}}, \tag{1.14}$$

where $\phi_k(x) := \tilde{\mathcal{L}}_{\mu_k}(x, y_k) + \mathbf{1}_\chi(x) + \frac{1}{2\mu_k}\|x - x_{k-1}\|_2^2$, $\tilde{\mathcal{L}}_{\mu_k}$ is defined in (1.10) and $\mathbf{1}_\chi$ is the indicator function of the closed convex set χ. With this new stopping condition, Rockafellar [14] was able to establish a *lower* bound $p(x_k) - p^* \geq -\mathcal{O}(\mu_k^{-1})$. Note that the stopping condition (1.14) is much stronger than (1.11) – in this paper we establish the lower bound using the weaker stopping condition (1.11).

First order methods for minimizing functions with Lipschitz continuous gradients [10, 11] (and also the non-smooth variants [3, 17]) can only guarantee convergence in function values; therefore, the subgradient condition (1.14) has to be re-stated in terms of function values in order to use a first-order algorithm to compute the iterates. This is impossible when the objective function is non-smooth. Therefore, one cannot establish operational level complexity results for a method that uses the gradient stopping condition (1.14) with first order methods. Next, consider the case where p is smooth, i.e. $\rho(\cdot) = 0$. Suppose $\chi = \mathbb{R}^n$, $\nabla\gamma$ is Lipschitz continuous with constant L_γ and $f(x) = Ax - b$. Then, it is easy to establish that $\nabla\phi_k$ is also Lipschitz continuous with Lipschitz constant $L_\phi = L_\gamma + \mu_k\sigma_{\max}^2(A) + \mu_k^{-1} = \mathcal{O}(\mu_k)$. Since $\phi_k(x_k) - \inf_{x\in\mathbb{R}^n} \phi_k(x) \leq \xi$ implies that $\|\nabla\phi_k(x_k)\|_2 \leq \sqrt{2L_\phi\xi}$, in order to ensure (1.14) one has to set $\xi \leq \frac{1}{2\sigma_{\max}^2(A)} \frac{\alpha_k}{\mu_k^2}$. Thus, the complexity of computing each iterate x_k satisfying (1.14) will be significantly higher than the complexity of computing x_k satisfying (1.11), which is the one used in the ALCC algorithm. Therefore, although Rockafellar's method using (1.14) has the same iteration complexity with ALCC algorithm, the operational level complexity of a first-order algorithm based on the gradient stopping criterion (1.14) will be significantly higher than the complexity of the ALCC algorithm where $\xi = \alpha_k$. In summary, Rockafellar [14] is only able to show an upper bound on sub-optimality of iterates for the stopping criterion (1.11) that leads to an efficient algorithm; whereas the subgradient stopping criterion (1.14) that results in a lower bound is not practical for a first-order algorithm.

In [6], Lan, Lu and Monteiro consider problems of the form

$$\min\{\langle c, x\rangle : \ Ax = b, \ x \in \mathcal{K}\}, \tag{1.15}$$

where $\mathcal{K}$ is a closed convex cone. They proposed computing an approximate solution for (1.15) by minimizing the Euclidean distance to the set of KKT points using Nesterov's accelerated proximal gradient algorithm (APG) [10, 11]. They show that at most $\mathcal{O}\left(\epsilon^{-1}\right)$ iterations of Nesterov's APG algorithm [10, 11] suffice to compute a point whose distance to the set of KKT points is at most $\epsilon > 0$. In [8], Lan and Monteiro proposed a first-order penalty method to solve the following more general problem

$$\min\{\gamma(x) : \ Ax - b \in \mathcal{K}, \ x \in \chi\}, \tag{1.16}$$

where γ is a convex function with Lipschitz continuous gradient, $\mathcal{K}$ is a closed, convex cone, χ is a simple convex compact set and $A \in \mathbb{R}^{m\times n}$. In order to solve (1.16), they used Nesterov's APG algorithm on the perturbed penalty problem

$$\min\{\gamma(x) + \xi\|x - x_0\|_2^2 + \frac{\mu}{2} \ d_{\mathcal{K}}(Ax - b)^2 : \ x \in \chi\},$$

where $x_0 \in \chi$, $d_{\mathcal{K}}$ is as defined in (1.7), and $\xi > 0$, $\mu > 0$ are fixed perturbation and penalty parameters. They showed that Nesterov's APG algorithm can compute a primal-dual solution $(\tilde{x}, \tilde{y}) \in \chi \times \mathcal{K}^*$ satisfying ϵ-perturbed KKT conditions

$$\langle \tilde{y},\ \Pi_{\mathcal{K}}(A\tilde{x} - b)\rangle = 0, \quad d_{\mathcal{K}}(A\tilde{x} - b) \leq \epsilon, \quad \nabla\gamma(\tilde{x}) - A^T\tilde{y} \in -\mathcal{N}_\chi(\tilde{x}) + \mathcal{B}(\epsilon), \tag{1.17}$$

using $\mathcal{O}\left(\epsilon^{-1}\log\left(\epsilon^{-1}\right)\right)$ projections onto $\mathcal{K}$ and χ, where $\mathcal{N}_\chi(\tilde{x}) := \{s \in \mathbb{R}^n :\ \langle s, x - \tilde{x}\rangle \leq 0,\ \forall x \in \chi\}$ and $\mathcal{B}(\epsilon) := \{x \in \mathbb{R}^n :\ \|x\|_2 \leq \epsilon\}$. Note that since ξ and μ are fixed, additional iterations of the Nesterov's APG algorithm will not improve the quality of the solution.

The optimization problem (1.16) is a special case of (1.1) with $\rho(\cdot) = 0$. Thus, ALCC can solve (1.16). We show that every limit point of the ALCC iterates are optimal for (1.16). Furthermore, for *any* $\epsilon > 0$, ALCC iterates are ϵ-optimal, and ϵ-feasible for (1.16) within $\mathcal{O}\left(\epsilon^{-1}\log\left(\epsilon^{-1}\right)\right)$ projections onto $\mathcal{K}$ and χ as is the case with the algorithm proposed in [8].

Lan and Monteiro [7] proposed an inexact augmented Lagrangian method to solve a special case of (1.1) with $\mathcal{K} = \{\mathbf{0}\}$ and $\rho(\cdot) = 0$; and showed that Nesterov's APG algorithm can compute a primal-dual solution $(\tilde{x}, \tilde{y}) \in \chi \times \mathbb{R}^m$ satisfying (1.17) using $\mathcal{O}\left(\epsilon^{-1}\left(\log\left(\epsilon^{-1}\right)\right)^{\frac{3}{4}}\log\log\left(\epsilon^{-1}\right)\right)$ projections onto χ and $\mathcal{K}$.

Aybat and Iyengar [2] proposed an inexact augmented Lanrangian algorithm (FALC) to solve the composite norm minimization problem

$$\min_{X \in \mathbb{R}^{m \times n}} \{\mu_1 \|\sigma(\mathcal{F}(X) - G)\|_\alpha + \mu_2 \|\mathcal{C}(X) - d\|_\beta + \gamma(X) :\ \mathcal{A}(X) - b \in \mathcal{Q}\}, \tag{1.18}$$

where the function $\sigma(\cdot)$ returns the singular values of its argument; α and $\beta \in \{1, 2, \infty\}$; $\mathcal{A}, \mathcal{C}, \mathcal{F}$ are linear operators such that either $\mathcal{C}$ or $\mathcal{F}$ is injective, and $\mathcal{A}$ is surjective; γ is a convex function with a Lipschitz continuous gradient and $\mathcal{Q}$ is a closed convex set. It was shown that any limit point of the FALC iterates is an optimal solution of the composite norm minimization problem (1.18); and for all $\epsilon > 0$, the FALC iterates are ϵ-feasible and ϵ-optimal after $\mathcal{O}\left(\log\left(\epsilon^{-1}\right)\right)$ FALC iterations, which require $\mathcal{O}\left(\epsilon^{-1}\right)$ shrinkage type operations and Euclidean projection onto the set $\mathcal{Q}$. The limitation of FALC is that it requires $\mathcal{A}$ to be a surjective mapping. Consider a feasible set of the form

$$\{x \in \mathbb{R}^n :\ A_1 x - b_1 \in \mathcal{K}_1,\ A_2 x - b_2 \in \mathcal{K}_2,\ x \in \chi\}, \tag{1.19}$$

where $\mathcal{K}_i$ is a closed convex cone, $A_i \in \mathbb{R}^{m_i \times n}$ and $b_i \in \mathbb{R}^{m_i}$ for $i = 1, 2$. The set in (1.19) can be reformulated as the feasible set in (1.1) by choosing $A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$ and $\mathcal{K} = \mathcal{K}_1 \times \mathcal{K}_2$, where $m = m_1 + m_2$. FALC can work with such a set only if A has linearly independent rows, i.e., $\mathbf{rank}(A) = m_1 + m_2$. This is a severe limitation for the practical problem. On the other hand, the ALCC algorithm works for the feasible sets of the form (1.19) without any additional assumption. Thus, ALCC can be used to solve much larger class of optimization problems.

In our opinion the ALCC algorithm proposed in this paper unifies all the previous work on fast first-order penalty and/or augmented Lagrangian algorithms for solving optimization problems that are special cases of (1.1). We do not impose any regularity conditions on the constraint matrix A and the projection step (1.2) is the natural extension of the gradient projection step. We believe that this unified treatment will spur further research in understanding the limits of performance of the first order algorithms for general conic problems.

2. Preliminaries.

In Section 2.1, first we briefly discuss a variant of Nesterov's APG algorithm [10, 11] to solve (1.1) without conic constraints. Next, we introduce a dual function for the conic problem in (1.1) and establish some of its properties in Section 2.2. The definitions and the results of Section 2.2 are extensions of the corresponding definitions and results in [12, 13], to the case where $\mathcal{K} \subset \mathbb{R}^m$ is a general closed, convex cone.

2.1. Accelerated Proximal Gradient (APG) algorithm.

In this section we state and briefly discuss the details of a particular implementation of Fast Iterative Shrinkage-Thresholding Algorithm [3] (FISTA), which extends Nesterov's accelerated proximal gradient algorithm [10, 11] for minimizing smooth convex functions over simple convex sets, to solve non-smooth convex minimization problems.

Algorithm APG$(\bar{\rho}, \bar{\gamma}, \chi, x_0, \text{STOP})$

$x_0^{(1)} \leftarrow x_0,\ x_1^{(2)} \leftarrow x_0,\ t_1 \leftarrow 1,\ \ell \leftarrow 0$
while STOP is **false do**
  $\ell \leftarrow \ell + 1$
  $x_\ell^{(1)} \leftarrow \operatorname{argmin}\left\{\bar{\rho}(x) + \left\langle \nabla\bar{\gamma}\left(x_\ell^{(2)}\right),\ x - x_\ell^{(2)}\right\rangle + \frac{L_{\bar{\gamma}}}{2}\ \|x - x_\ell^{(2)}\|_2^2 : x \in \chi\right\}$
  $t_{\ell+1} \leftarrow \left(1 + \sqrt{1 + 4\ t_\ell^2}\right)/2$
  $x_{\ell+1}^{(2)} \leftarrow x_\ell^{(1)} + \left(\frac{t_\ell - 1}{t_{\ell+1}}\right)\left(x_\ell^{(1)} - x_{\ell-1}^{(1)}\right)$
end while

Fig. 2.1: Accelerated Proximal Gradient Algorithm

FISTA computes an ϵ-optimal solution to $\min\{\bar{\rho}(x) + \bar{\gamma}(x) :\ x \in \mathbb{R}^n\}$ in $\mathcal{O}\left(\epsilon^{-\frac{1}{2}}\right)$ iterations, where $\bar{\rho} : \mathbb{R}^n \to \mathbb{R}$ and $\bar{\gamma} : \mathbb{R}^n \to \mathbb{R}$ are continuous convex functions such that $\nabla\bar{\gamma}$ is Lipschitz continuous on $\mathbb{R}^n$ with constant $L_{\bar{\gamma}}$. Tseng [17] showed that this rate result for FISTA also holds when $\bar{\rho} : \mathbb{R}^n \to (-\infty, +\infty]$ and $\bar{\gamma} : \mathbb{R}^n \to (-\infty, +\infty]$ are proper, lower semicontinuous, and convex functions such that $\mathbf{dom}\,\bar{\rho}$ is closed and $\nabla\bar{\gamma}$ is Lipschitz continuous on $\mathbb{R}^n$.

This extended version of FISTA is displayed in Figure 2.1 as APG algorithm. Hence, FISTA can solve constrained problems of the form

$$\min\{\bar{\rho}(x) + \bar{\gamma}(x) :\ x \in \chi\}, \tag{2.1}$$

where $\chi \subset \mathbb{R}^n$ is a simple closed convex set.

The APG algorithm displayed in Figure 2.1 takes as input the functions $\bar{\rho}$ and $\bar{\gamma}$, the simple closed convex set $\chi \subset \mathbb{R}^n$, an initial iterate $x^{(0)} \in \chi$ and a stopping criterion STOP. Lemma 2.1 gives the iteration complexity of the APG algorithm.

LEMMA 2.1. *Let $\bar{\rho}$ and $\bar{\gamma}$ be a proper, closed, convex functions such that $\mathbf{dom}\,\bar{\rho}$ is closed and $\nabla\bar{\gamma}$ is Lipschitz continuous on $\mathbb{R}^n$ with constant $L_{\bar{\gamma}}$. Fix $\epsilon > 0$ and let $\{x_\ell^{(1)}, x_\ell^{(2)}\}$ denote the sequence of iterates computed by the APG algorithm when* STOP *is disabled. Then $\bar{\rho}\big(x_\ell^{(1)}\big) + \bar{\gamma}\big(x_\ell^{(1)}\big) \leq \min\{\bar{\rho}(x) + \bar{\gamma}(x) :\ x \in \chi\} + \epsilon$ whenever $\ell \geq \sqrt{\frac{2L_{\bar{\gamma}}}{\epsilon}}\ \|x^* - x_0\|_2 - 1$, where $x^* \in \operatorname{argmin}\{\bar{\rho}(x) + \bar{\gamma}(x) :\ x \in \chi\}$.*

Proof. See Corollary 3 in [17] and Theorem 4.4 in [3] for the details of proof.□

2.2. A dual function for conic convex programs and its properties. For all $\mu \geq 0$, optimization problem (P) in (1.1) is equivalent to

$$\min\left\{p(x) + \frac{\mu}{2}\|Ax - s - b\|_2^2 : Ax - s = b,\ x \in \chi,\ s \in \mathcal{K}\right\}. \tag{2.2}$$

Let $y \in \mathbb{R}^m$ denote a Lagrangian dual variable corresponding to the equality constraint in (2.2), and let

$$\mathcal{L}_\mu(x, y) := \min_{s \in \mathcal{K}}\left\{p(x) - \langle y, Ax - s - b\rangle + \frac{\mu}{2}\|Ax - s - b\|_2^2\right\} \tag{2.3}$$

denote the "penalty" Lagrangian function for (2.2) with $\mathbf{dom}\,\mathcal{L}_\mu = \chi \times \mathbb{R}^m$. For $\mu > 0$,

$$\begin{aligned}\mathcal{L}_\mu(x, y) &= p(x) + \frac{\mu}{2}\ \left(\min_{s \in \mathcal{K}}\left\|Ax - s - b - \frac{y}{\mu}\right\|_2^2 - \frac{\|y\|_2^2}{\mu^2}\right),\\ &= p(x) + \frac{\mu}{2}\ d_{\mathcal{K}}\left(Ax - b - \frac{y}{\mu}\right)^2 - \frac{\|y\|_2^2}{2\mu},\end{aligned} \tag{2.4}$$

where $d_{\mathcal{K}}(\cdot)$ is the distance function defined in (1.7). When $\mu = 0$, the definition in (2.3) implies that

$$\mathcal{L}_0(x, y) = \begin{cases} p(x) - \langle y, Ax - b\rangle, & y \in \mathcal{K}^*,\\ -\infty, & \text{otherwise.}\end{cases} \tag{2.5}$$

For $\mu \geq 0$, we define a dual function $g_\mu : \mathbb{R}^m \to \mathbb{R}$ for (1.1) such that

$$g_\mu(y) := \inf_{x \in \chi} \mathcal{L}_\mu(x, y). \tag{2.6}$$

Note that from (2.5) it follows that g_0 is the Lagrangian dual function of (P).

The definitions above and the results detailed below are immediate extensions of corresponding definitions and results in [12], given for $\mathcal{K} = \mathbb{R}^m_+$, to the case where $\mathcal{K}$ is a general closed convex cone. We state and prove the extensions here for the sake of completeness. These results are used in Section 3 to establish the convergence properties of ALCC iterate sequence.

LEMMA 2.2. *For all $\mu \geq 0$, $x \in \chi$ and $y \in \mathbb{R}^m$, $\mathcal{L}_\mu$ defined in (2.3) satisfies*

$$\mathcal{L}_\mu(x, y) = \inf_{u \in \mathbb{R}^m} \left\{ F_\mu(x, u) - \langle y, u \rangle \right\}, \tag{2.7}$$

where $F_\mu : \chi \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is defined as follows

$$F_\mu(x, u) := \begin{cases} p(x) + \frac{\mu}{2}\|u\|_2^2, & \text{if } Ax - b \in \mathcal{K} + u, \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.8}$$

Hence, $\mathcal{L}_\mu(x, y)$ is convex in $x \in \chi$ and concave in $y \in \mathbb{R}^m$, and $g_\mu(y)$ defined in (2.6) is concave in $y \in \mathbb{R}^m$.

Proof. The representation in (2.7) trivially follows from the definition of F_μ in (2.8). For a fixed $x \in \chi$, (2.3) implies that $\mathcal{L}_\mu(x, y)$ is the infimum of affine functions of y, hence $\mathcal{L}_\mu(x, y)$ is concave in y. Hence, g_μ defined in (2.6) is the infimum of concave functions; therefore, it is also concave. For a fixed $y \in \mathbb{R}^m$, when $\mu > 0$, convexity of $\mathcal{L}_\mu(x, y)$ in x follows from (2.4) and the fact that $p(\cdot)$ and $d_\mathcal{K}(\cdot)$ are convex functions; otherwise, when $\mu = 0$, it trivially follows from (2.5). □

LEMMA 2.3. *Let $g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ be a proper closed convex function. For $\mu > 0$, let*

$$\psi_\mu(y) = \min_{z \in \mathbb{R}^m} \left\{ g(z) + \frac{1}{2\mu}\|z - y\|_2^2 \right\}, \quad \pi_\mu(y) = \operatorname*{argmin}_{z \in \mathbb{R}^m} \left\{ g(z) + \frac{1}{2\mu}\|z - y\|_2^2 \right\}$$

denote the Moreau regularization of and the proximal map corresponding to g, respectively. Then, for all $y_1, y_2 \in \mathbb{R}^m$,

$$\|\pi_\mu(y_1) - \pi_\mu(y_2)\|_2^2 + \|\pi_\mu^c(y_1) - \pi_\mu^c(y_2)\|_2^2 \leq \|y_1 - y_2\|_2^2, \tag{2.9}$$

where $\pi_\mu^c(y) := y - \pi_\mu(y)$ for all $z \in \mathbb{R}^m$. Moreover, $\psi_\mu : \mathbb{R}^m \to \mathbb{R}$ is an everywhere finite, differentiable convex function such that

$$\nabla \psi_\mu(y) = \frac{1}{\mu}\ (y - \pi_\mu(y)) = \frac{1}{\mu}\ \pi_\mu^c(y), \tag{2.10}$$

is Lipschitz continuous with constant $\frac{1}{\mu}$.

Proof. The proof of (2.9) is given in [15] and the rest of the claims including (2.10) are shown in [5]. □

THEOREM 2.4. *Suppose Assumption 1.1 holds. Then, for any $\mu > 0$, g_μ is an everywhere finite, continuously differentiable concave function and g_μ achieves its maximum value at any KKT point. Moreover,*

$$g_\mu(y) = \max_{z \in \mathbb{R}^m} \left\{ g_0(z) - \frac{1}{2\mu}\|z - y\|_2^2 \right\}, \tag{2.11}$$

and

$$\nabla g_\mu(y) = -\frac{1}{\mu}(y - \pi_\mu(y)), \tag{2.12}$$

is Lipschitz continuous with Lipschitz constant equal to $\frac{1}{\mu}$, where $\pi_\mu(y) \in \mathcal{K}^*$ denotes the unique maximizer in (2.11).

Proof. Fix $\mu \geq 0$, define

$$h_\mu(u) := \inf_{x\in\chi} F_\mu(x,u). \tag{2.13}$$

Note that $F_\mu(x,u) = p(x) + \frac{\mu}{2}\|u\|_2^2 + \mathbf{1}_{\mathcal{K}}(Ax-b-u)$, where $\mathbf{1}_{\mathcal{K}}(\cdot)$ denotes the indicator function of the set $\mathcal{K}$; therefore, $F_\mu(x,u)$ is convex in (x,u). Since F_μ is convex in (x,u), χ is a convex set and $h_\mu(0) = \inf_{x\in\chi}\{p(x) + \mathbf{1}_{\mathcal{K}}(Ax-b)\} = p^* > -\infty$, it follows that h_μ is a convex function such that $h_\mu(\cdot) > -\infty$ [4]. From the definition of F_μ, it follows that for all $u \in \mathbb{R}^m$,

$$h_\mu(u) = h_0(u) + \mu\ \omega(u),$$

where $\omega(u) := \frac{1}{2}\|u\|_2^2$. Substituting (2.7) in (2.6), for all $\mu \geq 0$, we get

$$g_\mu(y) = \inf_{u\in\mathbb{R}^m}\{h_\mu(u) - \langle y,u\rangle\} = -h_\mu^*(y),$$

where h_μ^* denotes the conjugate of the convex function h_μ.

Fix $\mu > 0$, since h_μ is a sum of two convex functions, it follows from Theorem 16.4 in [16] that

$$g_\mu(y) = -(h_0+\mu\omega)^*(y) = -\min_{z\in\mathbb{R}^m}\left\{h_0^*(z) + \mu\ \omega^*\left(\frac{y-z}{\mu}\right)\right\}. \tag{2.14}$$

Since $h_0^* = -g_0$ and $\omega^* = \omega$, the result (2.11) immediately follows from (2.14).

Note that (2.11) shows that $-g_\mu$ is the Moreau regularization of $-g_0$. Therefore, Lemma 2.3 and (2.11) imply that g_μ is everywhere finite, differentiable concave function such that ∇g_μ is given in (2.12).

Let y^* be a KKT point of (1.1). Note that $\pi_\mu(y^*) = y^*$. Hence $\nabla g_\mu(y^*) = \mathbf{0}$. Concavity of g_μ implies that $y^* \in \operatorname{argmax} g_\mu(y)$ for any KKT point y^*. □

THEOREM 2.5. *Fix $\mu > 0$ and $\bar{y} \in \mathbb{R}^m$. Suppose $\bar{x} \in \chi$ is an ξ-optimal solution to $\min_{x\in\chi} L_\mu(x,\bar{y})$, i.e. $L_\mu(\bar{x},\bar{y}) \leq \min\{L_\mu(x,\bar{y}) :\ x\in\chi\} + \xi = g_\mu(\bar{y}) + \xi$. Then*

$$\mu\ \|\nabla_y\mathcal{L}_\mu(\bar{x},\bar{y}) - \nabla g_\mu(\bar{y})\|_2^2 \leq 2\xi. \tag{2.15}$$

Proof. For $\mu > 0$, g_μ is concave and ∇g_μ is Lipschitz continuous with Lipschitz constant equal to $\frac{1}{\mu}$; therefore,

$$g_\mu(y) \geq g_\mu(\bar{y}) + \langle\nabla g_\mu(\bar{y}), y-\bar{y}\rangle - \frac{1}{2\mu}\|y-\bar{y}\|_2^2, \tag{2.16}$$

for all $y \in \mathbb{R}^m$. Moreover, since for every $x \in \chi$, $\mathcal{L}_\mu(x,y)$ is concave in y, it follows that for all $y \in \mathbb{R}^m$

$$\mathcal{L}_\mu(\bar{x},\bar{y}) + \langle\nabla_y\mathcal{L}_\mu(\bar{x},\bar{y}),\ y-\bar{y}\rangle \geq \mathcal{L}_\mu(\bar{x},y) \geq g_\mu(y). \tag{2.17}$$

Combining (2.16), (2.17) and the fact that $\bar{x}$ is ξ-optimal and y is arbitrary, we get

$$\xi \geq \sup_{y\in\mathbb{R}^m}\left\{\langle\nabla g_\mu(\bar{y}) - \nabla_y\mathcal{L}_\mu(\bar{x},\bar{y}),\ y-\bar{y}\rangle - \frac{1}{2\mu}\ \|y-\bar{y}\|_2^2\right\} = \frac{\mu}{2}\|\nabla g_\mu(\bar{y}) - \nabla_y\mathcal{L}_\mu(\bar{x},\bar{y})\|_2^2.$$

□

3. ALCC Algorithm.

In order to solve (P) given in (1.1), we inexactly solve the sequence of subproblems:

$$(SP_k):\ \min_{x\in\chi} P_k(x,y_k), \tag{3.1}$$

where

$$P_k(x,y) := \frac{1}{\mu_k}\ \mathcal{L}_{\mu_k}(x,y) = \frac{1}{\mu_k}\ p(x) + \frac{1}{2}\ d_{\mathcal{K}}\left(Ax - b - \frac{y}{\mu_k}\right)^2.$$

Algorithm ALCC $(x_0,\ \{\alpha_k,\ \eta_k,\ \mu_k\})$

$y_1 \leftarrow \mathbf{0}$, $k \leftarrow 1$
while $k \geq 1$ **do**
  $x_k \leftarrow \textsc{Oracle}(P_k, y_k, \alpha_k, \eta_k, \mu_k)$ /* *See Section 3.1 for* ORACLE */
  $y_{k+1} \leftarrow \mu_k \left[\Pi_{\mathcal{K}}\left(Ax_k - b - \frac{y_k}{\mu_k}\right) - \left(Ax_k - b - \frac{y_k}{\mu_k}\right)\right]$
  $k \leftarrow k+1$
end while

Fig. 3.1: Augmented Lagrangian Algorithm for Conic Convex Programming

For notational convenience, we define

$$f_k(x,y) := \frac{1}{2}\ d_{\mathcal{K}}\left(Ax - b - \frac{y}{\mu_k}\right)^2.$$

Therefore, $P_k(x,y) = \frac{1}{\mu_k}\ p(x) + f_k(x,y)$. The specific choice of penalty parameter and Lagrangian dual sequences, $\{\mu_k\}$ and $\{y_k\}$, are discussed later in this section.

LEMMA 3.1. *For all $k \geq 1$ and $y \in \mathbb{R}^m$, $f_k(x,y)$ is convex in x. Moreover,*

$$\nabla_x f_k(x,y) = A^T\left(Ax - b - \frac{y}{\mu_k} - \Pi_{\mathcal{K}}\left(Ax - b - \frac{y}{\mu_k}\right)\right), \tag{3.2}$$

and $\nabla_x f_k(x,y)$ is Lipschitz continuous in x with constant $L = \sigma^2_{\max}(A)$.

Proof. See appendix for the proof. □

The ALCC algorithm is displayed in Figure 3.1. The inputs to ALCC are an initial point $x_0 \in \chi$ and a parameter sequence $\{\alpha_k,\ \eta_k,\ \mu_k\}$ such that

$$\alpha_k \searrow 0, \quad \eta_k \searrow 0, \quad 0 < \mu_k \nearrow \infty. \tag{3.3}$$

3.1. Oracle. The subroutine $\textsc{Oracle}(P, \bar{y}, \alpha, \eta, \mu)$ returns $\bar{x} \in \chi$ such that $\bar{x}$ satisfies one of the following two conditions:

$$0 \leq P(\bar{x},\bar{y}) - \inf_{x\in\chi} P(x,\bar{y}) \leq \frac{\alpha}{\mu}, \tag{3.4}$$

$$\exists q \in \partial_x P(\bar{x},\bar{y}) + \partial_x \mathbf{1}_\chi(\bar{x}) \ \text{ s.t. } \|q\|_2 \leq \frac{\eta}{\mu}, \tag{3.5}$$

where $\mathbf{1}_\chi(\cdot)$ denotes the indicator function of the set χ.

Let $\bar{\rho}_k(x) := \frac{1}{\mu_k}\ \rho(x)$ and $\bar{\gamma}_k(x) := \frac{1}{\mu_k}\ \gamma(x) + f_k(x, y_k)$. Then $\nabla\bar{\gamma}_k$ exists and is Lipschitz continuous with Lipschitz constant

$$L_{\bar{\gamma}_k} := \frac{1}{\mu_k}\ L_\gamma + \sigma^2_{\max}(A). \tag{3.6}$$

Let

$$\chi \supset \chi^*_k := \operatorname*{argmin}_{x\in\chi} P_k(x, y_k) \tag{3.7}$$

denote the set of optimal solutions to (SP_k). Then, Lemma 2.1 guarantees that the APG algorithm with the initial iterate $x_{k-1} \in \chi$ requires at most

$$\ell_{\max}(k) := \sqrt{\frac{2\mu_k L_{\bar{\gamma}_k}}{\alpha_k}}\ d_{\chi^*_k}(x_{k-1}) \tag{3.8}$$

iterations to compute $\frac{\alpha_k}{\mu_k}$-optimal solution to the k-th subproblem (SP_k) in (3.1). Thus, setting the stopping criterion STOP $= \{l \geq \ell_{\max}(k)\}$ ensures that the output of the APG algorithm satisfies (3.4). Thus, we have shown that there exists a subroutine $\textsc{Oracle}(P_k, y_k, \alpha_k, \eta_k, \mu_k)$ that can compute x_k satisfying either (3.4) or (3.5). As indicated earlier, the computational complexity of each iteration in the APG algorithm is dominated by the complexity of computing the solution to (1.2).

3.2. Convergence properties of ALCC algorithm. In this section we investigate the convergence rate of ALCC algorithm.

LEMMA 3.2. *Let $\mathcal{K} \subset \mathbb{R}^n$ denote a closed, convex cone and $\bar{x} \in \mathbb{R}^n$. Then $\bar{x} - \Pi_{\mathcal{K}}(\bar{x}) \in -\mathcal{K}^*$ and $\langle \bar{x} - \Pi_{\mathcal{K}}(\bar{x}),\ \Pi_{\mathcal{K}}(\bar{x})\rangle = 0$, where $\mathcal{K}^* = \{s \in \mathbb{R}^n :\ \langle s, x\rangle \geq 0\ \forall x \in \mathcal{K}\}$. Finally, if $x \in -\mathcal{K}^*$, then $\Pi_{\mathcal{K}}(x) = \mathbf{0}$.*

Proof. See appendix for the proof. □

From Lemma 3.2, it follows that the dual variable y_{k+1} computed in Line 4 of ALCC algorithm satisfies $y_{k+1} \in \mathcal{K}^*$. Also note that for all $k \geq 1$,

$$y_{k+1} = y_k + \mu_k \nabla_y \mathcal{L}_{\mu_k}(x_k, y_k). \tag{3.9}$$

Next, we establish that the sequence of dual variables $\{y_k\}$ generated by ALCC algorithm is bounded for an appropriately chosen parameter sequence.

LEMMA 3.3. *Let $\{x_k, y_k\} \in \chi \times \mathcal{K}^*$ be the sequence of primal-dual ALCC iterates for a given input parameter sequence $\{\alpha_k,\ \eta_k,\ \mu_k\}$ satisfying (3.3). Then, for all $k \geq 1$,*

$$0 \leq \mathcal{L}_{\mu_k}(x_k, y_k) - g_{\mu_k}(y_k) \leq \xi_k, \tag{3.10}$$

where

$$\xi_k = \max\{\alpha_k,\ \eta_k\ d_{\chi_k^*}(x_k)\}, \tag{3.11}$$

and $\chi_k^* \subset \chi$ is defined in (3.7).

Proof. Fix $k \geq 1$. Suppose $x_k = \textsc{Oracle}(P_k, y_k, \alpha_k, \eta_k, \mu_k)$ satisfies (3.4). Then we have

$$P_k(x_k, y_k) \leq \inf_{x \in \chi} P_k(x, y_k) + \frac{\alpha_k}{\mu_k} = \frac{g_{\mu_k}(y_k) + \alpha_k}{\mu_k}. \tag{3.12}$$

Suppose instead that $x_k = \textsc{Oracle}(P_k, y_k, \alpha_k, \eta_k, \mu_k)$ satisfies (3.5). Then, there exists $q_k \in \partial_x P_k(x_k, y_k) + \partial \mathbf{1}_{\chi}(x_k)$ such that $\|q_k\|_2 \leq \frac{\eta_k}{\mu_k}$. Since $P_k(x, y_k) + \mathbf{1}_{\chi}(x)$ is convex in x, it follows that

$$P_k(x_k, y_k) \leq \inf_{\bar{x} \in \chi_k^*} P_k(\bar{x}, y_k) + \langle q_k, x_k - \bar{x}\rangle \leq \frac{g_{\mu_k}(y_k) + \eta_k\ d_{\chi_k^*}(x_k)}{\mu_k}. \tag{3.13}$$

Since $P_k(x, y) = \frac{1}{\mu_k} \mathcal{L}_{\mu_k}(x, y)$, the desired result follows from (3.12) and (3.13). □

The following result was originally established in [13] for $\mathcal{K} = \mathbb{R}^m_+$. We state and prove the extension to general convex cones for completeness.

THEOREM 3.4. *Suppose $B := \sum_{k=1}^{\infty} \sqrt{2\ \xi_k \mu_k} < \infty$, where ξ_k is defined in (3.11). Then, for all $k \geq 1$, $\|y_k\|_2 \leq B + \|y^*\|_2$ where y^* is any KKT point of (P).*

Proof. Lemma 3.3 and Theorem 2.5 imply that $\sqrt{2\ \xi_k \mu_k} \geq \|\mu_k \nabla_y \mathcal{L}_{\mu_k}(x_k, y_k) - \mu_k \nabla g_{\mu_k}(y_k)\|_2$. Next, adding and subtracting y_k, and using (2.12) and (3.9), we get

$$\sqrt{2\ \xi_k \mu_k} \geq \|\mu_k \nabla_y \mathcal{L}_{\mu_k}(x_k, y_k) + y_k - (y_k + \mu_k \nabla g_{\mu_k}(y_k))\|_2 = \|y_{k+1} - \pi_{\mu_k}(y_k)\|_2, \tag{3.14}$$

Since $\sum_{k=1}^{\infty} \sqrt{2\ \xi_k \mu_k} < \infty$, it follows that $\xi_k \mu_k \to 0$. Thus, $\lim_{k \in \mathbb{Z}_+} \left(y_{k+1} - \pi_{\mu_k}(y_k)\right) = 0$.

Assumption 1.1 guarantees that a KKT point $y^* \in \mathcal{K}^*$ exists. Since $y^* \in \operatorname{argmax}_{y \in \mathbb{R}^m} g_0(y)$, Theorem 2.4 implies that $y^* \in \operatorname{argmax}_{y \in \mathbb{R}^m} g_{\mu_k}(y)$ for all $k \geq 1$. Therefore, $\nabla g_{\mu_k}(y^*) = 0$, and consequently, by (2.12), $y^* = \pi_{\mu_k}(y^*)$. Since π_{μ_k} is non-expansive, it follows that

$$\|\pi_{\mu_k}(y_k) - y^*\|_2 = \|\pi_{\mu_k}(y_k) - \pi_{\mu_k}(y^*)\|_2 \leq \|y_k - y^*\|_2.$$

Hence,

$$\begin{aligned}
\|y_{k+1} - y^*\|_2 &\leq \|y_{k+1} - \pi_{\mu_k}(y_k)\|_2 + \|\pi_{\mu_k}(y_k) - y^*\|_2, \\
&\leq \|y_{k+1} - \pi_{\mu_k}(y_k)\|_2 + \|y_k - y^*\|_2, \\
&\leq \sqrt{2\ \xi_k \mu_k} + \|y_k - y^*\|_2.
\end{aligned} \tag{3.15}$$

Since $y_1 = \mathbf{0}$, the desired result is obtained by summing the above inequality over k. ◻

In the rest of this section we investigate the convergence properties of ALCC for the multiplier sequence $\{\alpha_k, \eta_k, \mu_k\}$ defined as follows

$$\mu_k = \beta^k \ \mu_0, \quad \alpha_k = \frac{1}{k^{2(1+c)} \ \beta^k} \ \alpha_0, \quad \eta_k = \frac{1}{k^{2(1+c)} \ \beta^k} \ \eta_0, \tag{3.16}$$

for all $k \geq 1$, where $\beta > 1$, c, α_0, η_0 and μ_0 are all strictly positive. Thus, $\alpha_k \searrow 0$, $\eta_k \searrow 0$ and $\mu_k \nearrow \infty$.

Let $\infty > \Delta_\chi := \max_{x\in\chi} \max_{x'\in\chi} \|x - x'\|_2$ denote the diameter of the compact set χ. Clearly, $d_{\chi_k^*}(x_k) \leq \Delta_\chi$ for all $k \geq 1$, where $\chi_k^* \subset \chi$ is defined in (3.7). Hence, from the definition of ξ_k in (3.11), it follows that

$$\sqrt{\xi_k \mu_k} \leq \frac{1}{k^{1+c}} \ \sqrt{\mu_0 \max\{\alpha_0, \ \eta_0 \Delta_\chi\}}, \quad \forall k \geq 1, \tag{3.17}$$

and $\sum_{k=1}^\infty \sqrt{\xi_k \mu_k} < \infty$ as required by Theorem 3.4. First, we lower bound the sub-optimality as a function of primal infeasibility of the iterates.

THEOREM 3.5. *Let $\{x_k, y_k\} \in \chi \times \mathcal{K}^*$ be the sequence of primal-dual ALCC iterates corresponding to a parameter sequence $\{\alpha_k, \eta_k, \mu_k\}$ satisfying* (3.3)*. Then*

$$p(x_k) - p^* \geq -\|y^*\|_2 \ d_\mathcal{K}\left(Ax_k - b - \frac{y_k}{\mu_k}\right) + \frac{1}{\mu_k} \langle y_k, y^* \rangle,$$

where $y^* \in \mathcal{K}^*$ denotes any KKT point of (P) and p^* denotes the optimal value of (P) given in (1.1).

Proof. The dual function $g_0(y) = -\infty$ when $y \notin \mathcal{K}^*$; and for all $y \in \mathcal{K}^*$, the dual function g_0 of (P) can be equivalently written as

$$\begin{aligned} g_0(y) &= \langle b, y \rangle + \inf_{x\in\mathbb{R}^n} \left\{ p(x) + \mathbf{1}_\chi(x) - \left\langle A^T y, x \right\rangle \right\}, \\ &= \langle b, y \rangle - (p + \mathbf{1}_\chi)^*(A^T y). \end{aligned}$$

Hence, the dual of (P) is

$$(D): \quad \max_{y\in\mathcal{K}^*} \langle b, y \rangle - (p + \mathbf{1}_\chi)^*(A^T y). \tag{3.18}$$

Any KKT point $y^* \in \mathcal{K}^*$ is an optimal solution of (3.18). Let $b_k := b + \frac{y_k}{\mu_k}$ for all $k \geq 1$. For $\kappa > 0$, define

$$\begin{aligned} (\mathcal{P}_k): \quad & \min_{x\in\chi} \{p(x) + \kappa \ d_\mathcal{K}(Ax - b_k)\}, \\ &= \min_{x\in\mathbb{R}^n, s\in\mathcal{K}} \{p(x) + \mathbf{1}_\chi(x) + \kappa \ \|Ax - b_k - s\|_2\}, \\ &= \max_{\|w\|_2\leq\kappa} \ \min_{x\in\mathbb{R}^n, s\in\mathcal{K}} \{p(x) + \mathbf{1}_\chi(x) + \langle w, \ Ax - b_k - s \rangle\}, \\ &= \max_{\|w\|_2\leq\kappa} \left\{ -\langle b_k, w \rangle + \inf_{s\in\mathcal{K}} \langle -w, s \rangle - \sup_{x\in\mathbb{R}^n} \left\{ \left\langle -A^T w, x \right\rangle - (p(x) + \mathbf{1}_\chi(x)) \right\} \right\}. \end{aligned}$$

Since $\inf_{s\in\mathcal{K}} \langle -w, s \rangle > -\infty$, only if $-w \in \mathcal{K}^*$; by setting $y = -w$, we obtain the following dual problem $(\mathcal{D}_k)$ of $(\mathcal{P}_k)$:

$$(D_k): \quad \max_{\|y\|_2\leq\tau, \ y\in\mathcal{K}^*} \left\{ \langle b_k, y \rangle - (p + \mathbf{1}_\chi)^*(A^T y) \right\}.$$

Since $y^* \in \mathcal{K}^*$ is feasible to $(\mathcal{D}_k)$ for $\kappa = \|y^*\|_2$, and $x_k \in \chi$ is feasible to (P_k), weak duality implies that

$$p(x_k) + \|y^*\|_2 \ d_\mathcal{K}(Ax_k - b_k) \geq \langle b, y^* \rangle - (p + \mathbf{1}_\chi)^*(A^T y^*) + \frac{1}{\mu_k} \langle y_k, y^* \rangle = p^* + \frac{1}{\mu_k} \langle y_k, y^* \rangle,$$

where the equality follows from strong duality between (P) and (D). ◻

Next, we upper bound the suboptimality.

THEOREM 3.6. *Let $\{x_k, y_k\} \in \chi \times \mathcal{K}^*$ be the sequence of primal-dual ALCC iterates corresponding to a parameter sequence $\{\alpha_k, \eta_k, \mu_k\}$ satisfying* (3.3)*. Let p^* denote the optimal value of* (P)*. Then*

$$P_k(x_k, y_k) - \frac{1}{\mu_k}\ p^* \leq \frac{1}{\mu_k}\ \xi_k^* + \frac{1}{2\mu_k^2}\ \|y_k\|_2^2, \tag{3.19}$$

where $\xi_k^* = \max\{\alpha_k, \eta_k\ d_{\chi^*}(x_k)\}$ and χ^* denote the set of optimal solutions to (P)*.*

Proof. Fix $k \geq 1$ and let $x^* \in \chi^*$. Suppose that $x_k = \textsc{Oracle}(P_k, y_k, \alpha_k, \eta_k, \mu_k)$ satisfies (3.4). Then, since $x^* \in \chi$, from (3.12), it follows that

$$P_k(x_k, y_k) \leq \inf_{x \in \chi} P_k(x, y_k) + \frac{\alpha_k}{\mu_k} \leq P_k(x^*, y_k) + \frac{\alpha_k}{\mu_k}. \tag{3.20}$$

Next, suppose that $x_k = \textsc{Oracle}(P_k, y_k, \alpha_k, \eta_k, \mu_k)$ satisfies (3.5). Then, since $P_k(x, y_k) + \mathbf{1}_\chi(x)$ is convex in x for all $k \geq 1$, it follows that

$$P_k(x_k, y_k) \leq P_k(x^*, y_k) + \langle q_k, x_k - x^* \rangle \leq P_k(x^*, y_k) + \frac{\eta_k\ \|x_k - x^*\|_2}{\mu_k}. \tag{3.21}$$

From (3.20) and (3.21), it follows that

$$P_k(x_k, y_k) - \frac{1}{\mu_k}\ p^* \leq \frac{1}{2}\ d_{\mathcal{K}}\left(Ax^* - b - \frac{y_k}{\mu_k}\right)^2 + \frac{\max\{\alpha_k,\ \eta_k\ \|x_k - x^*\|_2\}}{\mu_k}. \tag{3.22}$$

Since $Ax^* - b \in \mathcal{K}$, Lemma A.2 implies that $d_{\mathcal{K}}\left(Ax^* - b - \frac{y_k}{\mu_k}\right) \leq \frac{\|y_k\|_2}{\mu_k}$. Moreover, since $x^* \in \chi^*$ is arbitrary, from (3.22) it follows that

$$P_k(x_k, y_k) - \frac{1}{\mu_k}\ p^* \leq \frac{\|y_k\|_2^2}{2\mu_k} + \frac{\max\{\alpha_k,\ \eta_k\ \inf_{x^* \in \chi^*} \|x_k - x^*\|_2\}}{\mu_k}. \tag{3.23}$$

□

Note that since $f_k(\cdot) \geq 0$, we have $P_k(x_k, y_k) \geq \frac{1}{\mu_k}\ p(x_k)$ for all $k \geq 1$. Hence,

$$p(x_k) - p^* \leq \xi_k^* + \frac{1}{2\mu_k}\ \|y_k\|_2^2. \tag{3.24}$$

Now, we establish a bound on the infeasibility of the primal ALCC iterate sequence.

THEOREM 3.7. *Let $\{x_k, y_k\} \in \chi \times \mathcal{K}^*$ denote the sequence of primal-dual ALCC iterates for a parameter sequence $\{\alpha_k, \eta_k, \mu_k\}$ satisfying* (3.3) *and $y^* \in \mathcal{K}^*$ be a KKT point of* (P)*. Then*

$$0 \leq d_{\mathcal{K}}\left(Ax_k - b\right) \leq \frac{\|y_k\|_2 + \|y_{k+1} - y_k\|_2}{\mu_k} \tag{3.25}$$

for all $k \geq 1$, where $\xi_k^* = \max\{\alpha_k, \eta_k\ d_{\chi^*}(x_k)\}$ and χ^* denote the set of optimal solutions to (P)*.*

Proof. From Step 4 in ALCC algorithm, it follows that

$$\begin{aligned}\frac{y_{k+1} - y_k}{\mu_k} &= \Pi_{\mathcal{K}}\left(Ax_k - b - \frac{y_k}{\mu_k}\right) - (Ax_k - b),\\ &= \Pi_{\mathcal{K}}\left(Ax_k - b - \frac{y_k}{\mu_k}\right) - \Pi_{\mathcal{K}}(Ax_k - b) + \Pi_{\mathcal{K}}(Ax_k - b) - (Ax_k - b).\end{aligned}$$

Hence,

$$d_{\mathcal{K}}(Ax_k - b) \leq \frac{\|y_{k+1} - y_k\|_2}{\mu_k} + \left\|\Pi_{\mathcal{K}}\left(Ax_k - b - \frac{y_k}{\mu_k}\right) - \Pi_{\mathcal{K}}(Ax_k - b)\right\|_2.$$

The result now follows from the fact that $\Pi_{\mathcal{K}}$ is non-expansive. □

In the next theorem we establish the convergence rate of ALCC algorithm.

THEOREM 3.8. *Let $\{x_k, y_k\} \in \chi \times \mathcal{K}^*$ denote the sequence of primal-dual ALCC iterates for a parameter sequence $\{\alpha_k, \eta_k, \mu_k\}$ satisfying* (3.16). *Then for all $\epsilon > 0$, $d_{\mathcal{K}}(Ax_k - b) \leq \epsilon$ and $|p(x_k) - p^*| \leq \epsilon$ within $\mathcal{O}\left(\log\left(\epsilon^{-1}\right)\right)$* ORACLE *calls, which require solving at most $\mathcal{O}\left(\epsilon^{-1}\log\left(\epsilon^{-1}\right)\right)$ problems of the form* (1.2).

Proof. To simplify the notation, let $\alpha_0 = \eta_0 = \mu_0 = 1$, and, without loss of generality, assume that $1 \leq \mathcal{D}$, where $\mathcal{D} := \max_{x\in\chi} d_{\chi^*}(x) \leq \Delta_\chi < \infty$. Then, clearly $d_{\chi^*}(x_k) \leq \mathcal{D}$ for all $k \geq 1$.

First, (3.25) implies that

$$d_{\mathcal{K}}(Ax_k - b) \leq \frac{1}{\beta^k}\left(\|y_k\|_2 + \|y_{k+1} - y_k\|_2\right). \tag{3.26}$$

Moreover, from Step 4 of ALCC algorithm, it follows that

$$d_{\mathcal{K}}\left(Ax_k - b - \frac{y_k}{\mu_k}\right) \leq \frac{\|y_{k+1}\|_2}{\mu_k} = \frac{1}{\beta^k}\|y_{k+1}\|_2. \tag{3.27}$$

Now, Theorem 3.5, (3.24) and (3.27) together imply that

$$|p(x_k) - p^*| \leq \frac{1}{\beta^k}\max\left\{\|y^*\|_2\left(\|y_{k+1}\|_2 + \|y_k\|_2\right),\ \frac{\mathcal{D}}{k^{2(1+c)}} + \frac{\|y_k\|_2^2}{2}\right\} \tag{3.28}$$

Theorem 3.4 shows that $\{y_k\}$ is a bounded sequence. Hence, from (3.26) and (3.28), we have

$$d_{\mathcal{K}}(Ax_k - b) = \mathcal{O}\left(\frac{1}{\beta^k}\right), \qquad |p(x_k) - p^*| = \mathcal{O}\left(\frac{1}{\beta^k}\right). \tag{3.29}$$

Hence, (3.29) implies that for all $\epsilon > 0$, an ϵ-optimal and ϵ-feasible solution to (P) can be computed within $\mathcal{O}\left(\log\left(\epsilon^{-1}\right)\right)$ iterations of ALCC algorithm.

The values of $L_{\bar{\gamma}_k}$, α_k and μ_k are given respectively in (3.6) and (3.16). Substituting them in the expression for $\ell_{\max}(k)$ in (3.8) and using the fact that $d_{\chi_k^*}(x_{k-1}) \leq \Delta_\chi$, we obtain

$$\ell_{\max}(k) \leq \sqrt{\frac{2L_\gamma}{\beta^k} + 2\sigma_{\max}^2(A)\ d_{\chi_k^*}(x_{k-1})}\ \beta^k k^{1+c} = \mathcal{O}\left(\beta^k k^{1+c}\right). \tag{3.30}$$

Hence, (3.30) imply that at most $\mathcal{O}\left(\epsilon^{-1}\log(\epsilon^{-1})\right)$ problems of the form (1.2) are solved during $\mathcal{O}\left(\log\left(\epsilon^{-1}\right)\right)$ iterations of ALCC algorithm. Indeed, let $N_\epsilon \in \mathbb{Z}_+$ denote total number of problems of the form (1.2) solved to compute an ϵ-optimal and ϵ-feasible solution to (P). From (3.29) and (3.30), it follows that there exists $c_1 > 0$ and $c_2 > 0$ such that

$$N_\epsilon \leq \sum_{k=1}^{\log_\beta\left(\frac{c_1}{\epsilon}\right)} \ell_{\max}(k) \leq \sum_{k=1}^{\log_\beta\left(\frac{c_1}{\epsilon}\right)} c_2\beta^k k^{1+c} \leq \frac{\beta}{\beta-1}\left(\frac{c_1}{\epsilon} - 1\right)\left(\log_\beta\left(\frac{c_1}{\epsilon}\right)\right)^{1+c}.$$

□

COROLLARY 3.9. *Let $\{x_k, y_k\} \in \chi \times \mathcal{K}^*$ denote the sequence of primal-dual ALCC iterates for a parameter sequence $\{\alpha_k,\ \eta_k,\ \mu_k\}$ satisfying* (3.16). *Then $\lim_{k\in\mathbb{Z}_+} p(x_k) = p^*$ and $\lim_{k\in\mathbb{Z}_+} d_{\mathcal{K}}(Ax_k - b) = 0$. Moreover, for all $\mathcal{S} \subset \mathbb{Z}_+$ such that $\bar{x} = \lim_{k\in\mathcal{S}} x_k$, $\bar{x}$ is an optimal solution to (P).*

Proof. Since χ is compact, Bolzano–Weierstrass theorem implies that there exists a subsequence $\mathcal{S} \subset \mathbb{Z}_+$ such that $\bar{x} = \lim_{k\in\mathcal{S}} x_k$ exists. Moreover, taking the limit of both sides of (3.26) and (3.28), we have $\lim_{k\in\mathbb{Z}_+} d_{\mathcal{K}}(Ax_k - b) = 0$ and $\lim_{k\in\mathbb{Z}_+} p(x_k) = p^*$. Hence, $\lim_{k\in\mathcal{S}} d_{\mathcal{K}}(Ax_k - b) = 0$ and $\lim_{k\in\mathcal{S}} p(x_k) = p^*$. □

Note that even though $p(x_k) \to p^*$, the primal iterates themselves may not converge.

Rockafellar [13] proved that the dual iterate sequence $\{y_k\}$ computed via (1.11)–(1.12), converges to a KKT point of (1.9). We want to extend this result to the case where $\mathcal{K}$ is a general convex cone. The proof in [13] uses the fact that the penalty multiplier μ is fixed in (1.11)–(1.12) and it is not immediately clear how to extend this result to the setting with $\{\mu_k\}$ such that $\mu_k \to \infty$. In Theorem 3.10, we extend Rockafellar's result in [13] to arbitrary convex cones $\mathcal{K}$ when $f(x) = Ax - b$ and the penalty multipliers $\mu_k \to \infty$. After we independently proved Theorem 3.10, we became aware of an earlier work of Rockafellar [14] where he also extends the dual convergence result in [13] to the setting where $\{\mu_k\}$ is an increasing sequence. See Section 1.4 for a detailed discussion of our contribution in relation to this earlier work by Rockafellar.

THEOREM 3.10. *Let $\{x_k, y_k\} \in \chi \times \mathcal{K}^*$ denote the sequence of primal-dual ALCC iterates corresponding to a parameter sequence $\{\alpha_k, \eta_k, \mu_k\}$ satisfying* (3.16). *Then $\bar{y} := \lim_{k\in\mathbb{Z}_+} y_k$ exists and $\bar{y}$ is a KKT point of (P) in* (1.1).

Proof. It follows from (3.14) that for all $k \geq 1$ we have

$$\lim_{k\in\mathbb{Z}_+} \|y_{k+1} - \pi_{\mu_k}(y_k)\|_2 \leq \lim_{k\in\mathbb{Z}_+} \sqrt{2\xi_k\mu_k} = 0, \tag{3.31}$$

where ξ_k is defined in (3.11). Moreover, Theorem 3.4 shows that $\{y_k\}$ is a bounded sequence. Hence, (3.31) implies that $\{\pi_{\mu_k}(y_k)\}$ is also a bounded sequence.

From (2.11), it follows that $g_{\mu_k}(y_k) = g_0(\pi_{\mu_k}(y_k)) - \frac{1}{2\mu_k}\ \|\pi_{\mu_k}(y_k) - y_k\|_2^2$ and $g_{\mu_k}(y_k) \geq g_0(y^*) - \frac{1}{2\mu_k}\ \|y^* - y_k\|_2^2$ for any KKT point y^*. Since $g_0(y^*) = p^*$, we have that

$$g_0(\pi_{\mu_k}(y_k)) \geq p^* - \frac{1}{2\mu_k}\ \|y^* - y_k\|_2^2. \tag{3.32}$$

Since $\{y_k\}$ is bounded, taking the limit inferior of both sides of (3.32) we obtain

$$\liminf_{k\in\mathbb{Z}_+} g_0(\pi_{\mu_k}(y_k)) \geq p^* - \lim_{k\in\mathbb{Z}_+} \frac{1}{2\mu_k}\ \|y^* - y_k\|_2^2 = p^*. \tag{3.33}$$

Moreover, since $\pi_{\mu_k}(y_k) \in \mathcal{K}^*$ for all $k \geq 1$, weak duality implies that $\limsup_{k\in\mathbb{Z}_+} g_0(\pi_{\mu_k}(y_k)) \leq p^*$. Thus, using (3.33), we have that

$$\lim_{k\in\mathbb{Z}_+} g_0(\pi_{\mu_k}(y_k)) = p^*. \tag{3.34}$$

Since $\{\pi_{\mu_k}(y_k)\}$ is bounded, there exists $\mathcal{S} \subset \mathbb{Z}_+$ and $\bar{y} \in \mathcal{K}^*$ such that

$$\bar{y} := \lim_{k\in\mathcal{S}} \pi_{\mu_k}(y_k) = \lim_{k\in\mathcal{S}} y_{k+1}, \tag{3.35}$$

where the last equality follows from (3.31).

From (2.3) and (2.6), it follows that

$$g_0(y) = \inf_{x\in\chi, s\in\mathcal{K}} \{p(x) - \langle y, Ax - s - b\rangle\}.$$

Hence, $-g_0$ is a pointwise supremum of linear functions, which are always closed. Lemma 3.1.11 in [10] establishes that $-g_0$ is a closed convex function. Since a closed convex function is always lower semicontinous, we can conclude that $-g_0$ is lower semicontinuous, or equivalently, g_0 is an upper semicontinuous function. Hence, (3.34) and (3.35) imply that

$$p^* = \lim_{k\in\mathbb{Z}_+} g_0(\pi_{\mu_k}(y_k)) = \limsup_{k\in\mathcal{S}} g_0(\pi_{\mu_k}(y_k)) \leq g_0(\bar{y}) \leq p^*,$$

where the first inequality is due to upper semicontinuity of g_0 and the last one is due to weak duality and the fact that $\bar{y} \in \mathcal{K}^*$. Thus, we have

$$g_0(\bar{y}) = \lim_{k\in\mathbb{Z}_+} g_0(\pi_{\mu_k}(y_k)) = p^*, \tag{3.36}$$

which implies that $\bar{y} \in \mathcal{K}^*$ is a KKT point of (1.1).

Moreover, since (3.15) holds for any KKT point, we can substitute $\bar{y}$ for y^* in the expression. Thus, we have

$$\|y_\ell - \bar{y}\|_2 \leq \|y_k - \bar{y}\|_2 + \sum_{t\geq k} \sqrt{2\xi_t\mu_t}, \quad \forall \ell > k. \tag{3.37}$$

Fix $\epsilon > 0$. Since the sequence $\{\sqrt{\xi_k \mu_k}\}$ is summable, it follows that there exists $N_1 \in \mathbb{Z}_+$ such that $\sum_{t=k}^{\infty} \sqrt{2\xi_t \mu_t} \leq \frac{\epsilon}{2}$ for all $k > N_1$. Moreover, since the $\{y_k\}_{k \in \mathcal{S}}$ converges to $\bar{y}$, it follows that there exists $N_2 \in \mathcal{S}$ such that $N_2 \geq N_1$ and $\|y_{N_2} - \bar{y}\|_2 \leq \frac{\epsilon}{2}$. Hence, (3.37) implies that $\|y_\ell - \bar{y}\|_2 \leq \epsilon$ for all $\ell > N_2$. Therefore, $\lim_{k \in \mathbb{Z}_+} y_k = \bar{y}$. ∎

4. Conclusion. In this paper we build on previously known augmented Lagrangian algorithms for convex problems with standard inequality constraints [12, 13] to develop the ALCC algorithm that solves convex problems with conic constraints. In each iteration of the ALCC algorithm, a sequence of "penalty" Lagrangians—see (2.4)—are inexactly minimized over a "simple" closed convex set. We show that recent results on optimal first-order algorithms [3, 17] (see also [10, 11]), can be used to bound the number of basic operations needed in each iteration to inexactly minimize the "penalty" Lagrangian sub-problem. By carefully controlling the growth of the penalty parameter μ_k that controls the iteration complexity of ALCC algorithm, and the decay of parameter α_k that controls the suboptimality of each sub-problem, we show that ALCC algorithm is a theoretically efficient first-order, inexact augmented Lagrangian algorithm for structured non-smooth conic convex programming.

REFERENCES

[1] N. S. Aybat and G. Iyengar, *A first-order augmented Lagrangian method for compressed sensing*, SIAM Journal on Optimization, 22 (2012), pp. 429–459.

[2] ———, *Unified approach for minimizing composite norms*, forthcoming in Mathematical Programming Journal, Series A, (2012).

[3] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 183–202.

[4] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.

[5] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods*, Springer-Verlag, New York, 2001.

[6] G. Lan, Z. Lu, and R. Monteiro, *Primal-dual first-order methods with $\mathcal{O}(1/\epsilon)$ iteration complexity for cone programming*, Mathematical Programming, Series A, 126 (2011), p. 129.

[7] G. Lan and R. Monteiro, *Iteration-complexity of first-order augmented lagrangian methods for convex programming*, submitted to Mathematical Programming, Series A, (2009).

[8] ———, *Iteration-complexity of first-order penalty methods for convex programming*, forthcoming, Mathematical Programming, Series A, (2012).

[9] B. Martinet, *Perturbation des méthodes d'optimisation. applications*, RAIRO, Analyse numérique, 12 (1978), pp. 153–171.

[10] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Kluwer Academic Publishers, 2004.

[11] ———, *Smooth minimization of nonsmooth functions*, Mathematical Programming, Series A, 103 (2005), pp. 127–152.

[12] R. T. Rockafellar, *A dual approach to solving nonlinear programming problems by unconstrained optimization*, Mathematical Programming, 5 (1973), pp. 354–373.

[13] ———, *The multiplier method of Hestenes and Powell applied to convex programming*, Journal of Optimization and Applications, 12 (1973), pp. 555–562.

[14] ———, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Math. Oper. Res., 1 (1976), pp. 97–116.

[15] ———, *Monotone operators and the proximal point algorithm*, SIAM Journal on Control and Optimization, 14 (1976), pp. 877–898.

[16] ———, *Convex Analysis*, Princeton University Press, Princeton, 1997, c1970.

[17] P. Tseng, *On accelerated proximal gradient methods for convex-concave optimization*, submitted to SIAM Journal on Optimization, (2008).

Appendix A. Proofs of technical results.

Lemma A.1. *Let $f(\cdot) = \frac{1}{2} d_{\mathcal{K}}^2(\cdot)$. Then f is convex, and $\nabla f(y) = y - \Pi_{\mathcal{K}}(y)$ is Lipschitz continuous with Lipschitz constant equal to 1. Moreover, both $\Pi_{\mathcal{K}}(\cdot)$ and $\Pi_{\chi}^c(z) = z - \Pi_{\mathcal{K}}(z)$ are nonexpansive.*

Proof. The indicator function $\mathbf{1}_{\mathcal{K}}(\cdot)$ of a closed convex set $\mathcal{K}$ is a proper closed convex function, and

$$f(y) = \min_{z \in \mathbb{R}^m} \{\mathbf{1}_{\mathcal{K}}(\cdot)(z) + \frac{1}{2}\|z - y\|_2^2\} = \min_{z \in \mathcal{K}} \frac{1}{2}\|z - y\|_2^2,$$

is the Moreau regularization of the function $\mathbf{1}_{\mathcal{K}}(\cdot)$, and the projection operator $\Pi_{\mathcal{K}}(\cdot)$ is the corresponding Moreau proximal map. Therefore, all the results of this lemma follow from Lemma 2.3. ∎

Lemma A.2. *For all $y, y' \in \mathbb{R}^m$, $d_{\mathcal{K}}(y) \leq d_{\mathcal{K}}(y+y') + \|y'\|_2$.*

Proof.

$$\begin{aligned} d_{\mathcal{K}}(y) &= \|\Pi_{\mathcal{K}}(y) - y\|_2 = \|\Pi_{\mathcal{K}}(y) - y + \Pi_{\mathcal{K}}(y+y') - \Pi_{\mathcal{K}}(y+y') + y' - y' + y - y\|_2, \\ &\leq \|\Pi_{\mathcal{K}}(y+y') - (y+y')\|_2 + \|\Pi^c_{\mathcal{K}}(y+y') - \Pi^c_{\mathcal{K}}(y)\|_2, \\ &\leq d_{\mathcal{K}}(y+y') + \|y'\|_2, \end{aligned}$$

where the last inequality follows from the fact that $\Pi^c_{\chi}(x) = x - \Pi_{\mathcal{K}}(x)$ is nonexpansive. ◻

Proof of Lemma 3.1.

Proof. For all $y \in \mathbb{R}^m$, the convexity of $f_k(x, y)$ in x follows from Lemma A.1.

Moreover, Lemma A.1 and the chain rule, together imply (3.2). Now, fix $x', x'' \in \mathbb{R}^n$ and $\bar{y} \in \mathbb{R}^m$. Then (3.2) implies that

$$\begin{aligned} &\|\nabla_x f_k(x', \bar{y}) - \nabla_x f_k(x'', \bar{y})\|_2 \\ &= \left\| A^T \left[Ax' - b - \frac{\bar{y}}{\mu_k} - \Pi_{\mathcal{K}} \left(Ax' - b - \frac{\bar{y}}{\mu_k} \right) - \left(Ax'' - b - \frac{\bar{y}}{\mu_k} - \Pi_{\mathcal{K}} \left(Ax'' - b - \frac{\bar{y}}{\mu_k} \right) \right) \right] \right\|_2, \\ &\leq \sigma_{\max}(A) \|A(x' - x'')\|_2 \leq \sigma^2_{\max}(A) \ \|x' - x''\|_2, \end{aligned}$$

where the first inequality follows from the non-expansiveness of $\Pi^c_{\chi}(\cdot)$. ◻

Proof of Lemma 3.2.

Proof. $\Pi_{\mathcal{K}}(x) \in \operatorname{argmin}_{s \in \mathcal{K}} \|s - x\|_2^2$, if, and only if, $\langle \Pi_{\mathcal{K}}(x) - x,\ s - \Pi_{\mathcal{K}}(x) \rangle \geq 0$ for all $s \in \mathcal{K}$. Hence,

$$\langle \Pi_{\mathcal{K}}(x) - x,\ s \rangle \geq \langle \Pi_{\mathcal{K}}(x) - x,\ \Pi_{\mathcal{K}}(x) \rangle, \qquad \forall s \in \mathcal{K}. \tag{A.1}$$

Since the left hand side of (A.1) is bounded from below for all $s \in \mathcal{K}$, it follows that $\Pi_{\mathcal{K}}(x) - x \in \mathcal{K}^*$. Moreover, since $\Pi_{\mathcal{K}}(x) \in \mathcal{K}$, we have

$$0 = \min_{s \in \mathcal{K}} \langle \Pi_{\mathcal{K}}(x) - x,\ s \rangle \geq \langle \Pi_{\mathcal{K}}(x) - x,\ \Pi_{\mathcal{K}}(x) \rangle \geq 0.$$

This implies $\langle \Pi_{\mathcal{K}}(x) - x,\ \Pi_{\mathcal{K}}(x) \rangle = 0$.

Suppose $x \in -\mathcal{K}$. Clearly, $\langle \mathbf{0} - x,\ s - \mathbf{0} \rangle \geq 0$ for all $s \in \mathcal{K}$. Thus, it follows that $\Pi_{\mathcal{K}}(x) = \mathbf{0}$. ◻