

RESEARCH

Open Access



# An ADMM-based SQP method for separably smooth nonconvex optimization

Meixing Liu<sup>1</sup> and Jinbao Jian<sup>2\*</sup>

\*Correspondence:

[jianjb@gxu.edu.cn](mailto:jianjb@gxu.edu.cn)

<sup>2</sup>College of Science, Guangxi University for Nationalities, Nanning, P.R. China

Full list of author information is available at the end of the article

## Abstract

This work is about a splitting approach for solving separably smooth nonconvex linearly constrained optimization problems. Based on the ideas from two classical methods, namely the sequential quadratic programming (SQP) and the alternating direction method of multipliers (ADMM), we propose an ADMM-based SQP method. We focus on decomposing the quadratic programming (QP) subproblem of the primal problem into small-scale QP subproblems, which further embedded with Bregman distances can be solved effectively and followed by a dual ascent type update for the Lagrangian multipliers. Under suitable conditions as well as the crucial Kurdyka–Łojasiewicz property, we establish the global and strong convergence properties of the proposed method.

**Keywords:** Nonconvex optimization; SQP; ADMM; Kurdyka–Łojasiewicz property; Convergence

## 1 Introduction

Nonconvex optimization problems arise in a variety of applications ranging from the fields of signal and image processing, machine learning [1]. This class of problems is often structured and explicitly characterized with a separable objective, although they may be rather challenging to deal with.

In this paper, we consider the following nonconvex optimization problems with linear constraints and a separable objective function:

$$\begin{aligned} \min_{x,y} \quad & f(x) + g(y), \\ \text{s.t.} \quad & Ax + By = b, \end{aligned} \tag{1}$$

where  $f : R^{n_1} \rightarrow R$  and  $g : R^{n_2} \rightarrow R$  are continuously differentiable, but not necessarily convex, matrices  $A \in R^{m \times n_1}$ ,  $B \in R^{m \times n_2}$  and the vector  $b \in R^m$  are assumed to be given.

To solve problem (1) with separable structure, when  $f$  and  $g$  are convex, one simple but powerful method is the alternating direction method of multipliers (ADMM), which was originally proposed in [2, 3]. A survey of ADMM or its related variants has gained in popularity by many researchers, see, e.g., [4–9]. The standard iterative scheme of the

© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

classical ADMM for problem (1) acts as follows:

$$\begin{cases} x^{k+1} = \arg \min_{x \in R^{n_1}} \{ \mathcal{L}_\beta(x, y^k, \lambda^k) \}, & (2a) \\ y^{k+1} = \arg \min_{y \in R^{n_2}} \{ \mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \}, & (2b) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b), & (2c) \end{cases}$$

where  $\beta > 0$  is a penalty parameter and

$$\mathcal{L}_\beta(x, y, \lambda) = f(x) + g(y) - \lambda^T(Ax + By - b) + \frac{\beta}{2} \|Ax + By - b\|^2 \quad (3)$$

is the augmented Lagrangian function with a Lagrangian multiplier  $\lambda \in R^m$ .

In contrast to the developments of ADMM for the convex case outlined above, there are some works investigated on understanding the properties of splitting approaches for nonconvex problems, although a rigorous analysis is generally very difficult. Note that the recent works in [10–12] all dealt with the nonconvex problems by means of ADMM-type methods and established favorably crucial convergence results. In particular, a Bregman modification of ADMM on the problem with the sum of a smooth function and a nonconvex function in the objective was considered by Wang et al. [10]. Meanwhile, Li et al. [11] devised two types of splitting methods, in which there was only one subproblem also embedded with a Bregman distance. Moreover, Hong et al. [12] focused on solving the nonconvex consensus and sharing problems. We believe that it is very meaningful and important to further study characteristics of splitting approaches designed in the sense of nonconvexity and enlarge applicable spectrums due to the necessity in practice.

On the other hand, it is well known that the sequential quadratic programming (SQP) method, dated back earliest to [13], is one of the most efficient methods for solving smoothly constrained optimization problems, since it enjoys good theoretical properties and stable numerical performance with better approximation of the primal problem. For more than half a century, the SQP method has received rapid development and fruitful achievements, see, e.g., [14–21] and the references therein. Recently, Jian et al. [22] discussed a class of separably smooth nonconvex optimization problems with linear constraints and closed convex sets and presented an ADMM-SQP method. In this work, the QP subproblem of the primal problem was split into two smaller-scale QP subproblems, which were solved in a Jacobian manner. Moreover, an inexact Armijo line search was carried out to illustrate the descent property of the augmented Lagrangian function, and the global convergence was proved under proper conditions. As we know, the classical QP subproblem associated with problem (1) reads as follows:

$$\begin{aligned} \min_{x,y} \quad & \nabla f(x^k)^T(x - x^k) + \nabla g(y^k)^T(y - y^k) + \frac{1}{2} \|(x - x^k, y - y^k)\|_{H_k}^2, \\ \text{s.t.} \quad & Ax + By = b, \end{aligned} \quad (4)$$

where  $H_k \in R^{(n_1+n_2) \times (n_1+n_2)}$  is a symmetric approximation to the Hessian of the Lagrangian function, namely  $\mathcal{L}_0(x, y, \lambda)$  with  $\beta = 0$  in (3), for problem (1) with respect to variables  $(x, y)$ .

Consider that

$$\nabla_{(x,y)}^2 \mathcal{L}_0(x, y, \lambda) = \begin{pmatrix} \nabla^2 f(x) & O \\ O & \nabla^2 g(y) \end{pmatrix},$$

then it is reasonable to choose matrix  $H_k$  of the form

$$H_k = \begin{pmatrix} H_k^x & O \\ O & H_k^y \end{pmatrix}, \quad (5)$$

where  $H_k^x$  and  $H_k^y$  are the symmetric approximations of  $\nabla^2 f(x^k)$  and  $\nabla^2 g(y^k)$ , respectively. In view of this, QP subproblem (4) can be converted into the following structure:

$$\begin{aligned} \min_{x,y} \quad & \nabla f(x^k)^T (x - x^k) + \frac{1}{2} \|x - x^k\|_{H_k^x}^2 + \nabla g(y^k)^T (y - y^k) + \frac{1}{2} \|y - y^k\|_{H_k^y}^2 \\ \text{s.t.} \quad & Ax + By = b, \end{aligned} \quad (6)$$

where we find that the objective function is separable.

In this paper, motivated by the ideas of the splitting scheme applied to the QP subproblem in [22], and of the Bregman modification of ADMM in [10], we focus on QP subproblem (6) of the primal problem, and propose an ADMM-based SQP algorithm in the nonconvex sense. The resulting method makes use of the separable structure of QP subproblem (6) and decomposes it into two relatively small-scale QP subproblems in a Gauss–Seidel manner, which further are well equipped with additional Bregman distances and then are solved effectively. The main difference from the work [22] is that our proposed method is irrelevant with any line search and the convergence properties can be proved in terms of the potential function under some suitable conditions.

The remainder of this paper is structured as follows. In Sect. 2, the ADMM-based SQP algorithm is established as some elementary preliminaries are prepared. Section 3 presents the convergence properties of the proposed algorithm. Finally, we give the conclusions in Sect. 4.

**Notation.** Throughout this paper,  $R^n$  stands for the  $n$ -dimensional real Euclidean space,  $I$  is an identity matrix,  $\|\cdot\|$  is the Euclidean norm equipped with inner product  $\langle \cdot, \cdot \rangle$ . For any vector  $x$  and matrix  $H$ , we denote  $\|x\|_H^2 := x^T H x$ , where  $T$  is the transpose operation.  $H \succ 0$  means that the matrix  $H$  is positive definite (resp., positive semidefinite,  $H \succeq 0$ ), while  $H \succ G$  is used to denote  $H - G \succ 0$  (resp.,  $H - G \succeq 0$ ,  $H \succeq G$ ), and moreover the minimum eigenvalue of a matrix  $H$  is denoted by  $\sigma_H$ . For brevity, we additionally introduce the following notations:

$$w := (x, y, \lambda), \quad w^k := (x^k, y^k, \lambda^k), \quad \hat{w} := (x, y, \lambda, \hat{y}), \quad \hat{w}^k := (x^k, y^k, \lambda^k, y^{k-1}),$$

and the primal-dual errors

$$\Delta x^k := x^k - x^{k-1}, \quad \Delta y^k := y^k - y^{k-1}, \quad \Delta \lambda^k := \lambda^k - \lambda^{k-1}.$$

## 2 Preliminaries and ADMM-based SQP method

In this section, we provide some preliminaries that are useful in the sequel, and then describe the ADMM-based SQP method in detail.

The domain of function  $f$  is defined as  $\text{dom} f := \{x \in \mathbb{R}^n : f(x) < +\infty\}$ . For a subset  $S \subseteq \mathbb{R}^n$  and a point  $x \in \mathbb{R}^n$ , the distance from  $x$  to  $S$  is defined as  $d(x, S) := \inf\{\|x - y\| : y \in S\}$  and by convention  $d(x, S) = +\infty$  for all  $x$  when  $S = \emptyset$ .

**Definition 1** ([23], KL property) Let  $f$  be a proper lower semicontinuous function, and let  $\partial f$  be the basic subdifferential of  $f$  in  $\text{dom} f$ . Then

- (a) The function  $f$  is said to have the Kurdyka–Łojasiewicz (KL) property at  $x^* \in \text{dom} \partial f$  if there exist  $\eta \in (0, +\infty]$ , a neighborhood  $U$  of  $x^*$ , and a continuous and concave function  $\varphi : [0, \eta] \rightarrow \mathbb{R}_+$  such that
- (i)  $\varphi(0) = 0$ , and  $\varphi$  is continuously differentiable on  $(0, \eta)$  with  $\varphi' > 0$ ;
  - (ii) for all  $x$  in  $U \cap \{x \in \mathbb{R}^n : f(x^*) < f(x) < f(x^*) + \eta\}$ , the KL inequality holds

$$\varphi'(f(x) - f(x^*))d(0, \partial f(x)) \geq 1.$$

- (b) Let  $\Phi_\eta$  be the set of concave functions that satisfy (i); if  $f$  satisfies the KL property at each point of  $\text{dom} \partial f$ , then  $f$  is called a KL function.

**Lemma 1** ([24], Uniformized KL property) Let  $\Omega$  be a compact set, and let  $f$  be a proper lower semicontinuous function. Assume that  $f$  is constant on  $\Omega$  and satisfies the KL property at each point of  $\Omega$ . Then there exist  $\epsilon > 0, \eta > 0$ , and  $\varphi \in \Phi_\eta$  such that, for all  $\tilde{x} \in \Omega$  and for all  $x$ , in the following intersection

$$\{x \in \mathbb{R}^n : d(x, \Omega) < \epsilon\} \cap \{x \in \mathbb{R}^n : f(\tilde{x}) < f(x) < f(\tilde{x}) + \eta\},$$

one has

$$\varphi'(f(x) - f(\tilde{x}))d(0, \partial f(x)) \geq 1.$$

A semialgebraic set  $S \subseteq \mathbb{R}^n$  is a finite union of sets of the form

$$\{x \in \mathbb{R}^n : p_1(x) = \cdots = p_k(x) = 0, q_1(x) < 0, \dots, q_l(x) < 0\},$$

where  $p_1, \dots, p_k$  and  $q_1, \dots, q_l$  are real polynomial functions. A function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is semialgebraic if its graph is a semialgebraic subset of  $\mathbb{R}^{n+1}$ . Such a function satisfies the KL property, see, e.g., [23, 25, 26], with  $\varphi(s) = cs^{1-\theta}$  for some  $\theta \in [0, 1)$  and some  $c > 0$ . On the other hand, some important stability properties of semialgebraic functions can be found [27].

- finite sums and products of semialgebraic functions are semialgebraic;
- scalar products are semialgebraic;
- indicator functions of semialgebraic sets are semialgebraic;
- generalized inverse of semialgebraic mappings are semialgebraic;
- composition of semialgebraic functions or mappings are semialgebraic;
- functions of the type  $\mathbb{R}^n \ni x \rightarrow f(x) = \sup_{y \in C} g(x, y)$  (resp.  $\mathbb{R}^n \ni x \rightarrow f(x) = \inf_{y \in C} g(x, y)$ ) where  $g$  and  $C$  are semialgebraic.

For a continuously differentiable convex function  $f$  on  $R^n$ , the associated Bregman distance  $\Delta_f$  is defined as

$$\Delta_f(x^1, x^2) = f(x^1) - f(x^2) - \langle \nabla f(x^2), x^1 - x^2 \rangle$$

for any  $x^1, x^2 \in R^n$ . Let us now collect some important properties about Bregman distance [10].

- *Nonnegativity*:  $\Delta_f(x^1, x^2) \geq 0$ ,  $\Delta_f(x^1, x^1) = 0$  for all  $x^1, x^2$ .
- *Convexity*:  $\Delta_f(x^1, x^2)$  is convex in  $x^1$ , but not necessarily in  $x^2$ .
- *Strong convexity*: If  $f$  is  $\sigma_f$ -strongly convex, then  $\Delta_f(x^1, x^2) \geq \frac{\sigma_f}{2} \|x^1 - x^2\|^2$  for all  $x^1, x^2$ .

For the current primal-dual iterate  $(x^k, y^k, \lambda^k) \in R^{n_1} \times R^{n_2} \times R^m$ , we apply the splitting idea of the classical ADMM to the structured QP subproblem (6) so that variables  $x$  and  $y$  are updated alternatively at each iteration with regularized Bregman distances  $\Delta_\phi(\cdot, x^k)$  and  $\Delta_\psi(\cdot, y^k)$ , respectively, and then followed by the update of the Lagrangian multiplier  $\lambda$ , such a procedure can be formulated as follows:

$$\begin{cases} x^{k+1} = \arg \min_{x \in R^{n_1}} \left\{ \nabla f(x^k)^T (x - x^k) + \frac{1}{2} \|x - x^k\|_{H_k^x}^2 - (\lambda^k)^T (Ax + By^k - b) \right. \\ \quad \left. + \frac{\beta}{2} \|Ax + By^k - b\|^2 + \Delta_\phi(x, x^k) \right\}, \end{cases} \quad (7a)$$

$$\begin{cases} y^{k+1} = \arg \min_{y \in R^{n_2}} \left\{ \nabla g(y^k)^T (y - y^k) + \frac{1}{2} \|y - y^k\|_{H_k^y}^2 - (\lambda^k)^T (Ax^{k+1} + By - b) \right. \\ \quad \left. + \frac{\beta}{2} \|Ax^{k+1} + By - b\|^2 + \Delta_\psi(y, y^k) \right\}, \end{cases} \quad (7b)$$

$$\lambda^{k+1} = \lambda^k - \beta (Ax^{k+1} + By^{k+1} - b), \quad (7c)$$

where  $\phi$  and  $\psi$  are continuously differentiable and strongly convex functions with modulus  $\sigma_\phi, \sigma_\psi$  on  $R^{n_1}$  and  $R^{n_2}$ , respectively. Notice that the objective functions of (7a) and (7b) are strictly convex if additionally we use the strong convexity of  $\phi$  and  $\psi$  as well as the conditions that  $H_k^x + \beta A^T A + \sigma_\phi I_{n_1} \succ 0$  and  $H_k^y + \beta B^T B + \sigma_\psi I_{n_2} \succ 0$ .

Invoking the first-order optimality conditions of (7a) and (7b), we have

$$\begin{aligned} \nabla f(x^k) + H_k^x \Delta x^{k+1} - A^T \lambda^k + \beta A^T (Ax^{k+1} + By^k - b) \\ + \nabla \phi(x^{k+1}) - \nabla \phi(x^k) = 0, \end{aligned} \quad (8a)$$

$$\begin{aligned} \nabla g(y^k) + H_k^y \Delta y^{k+1} - B^T \lambda^k + \beta B^T (Ax^{k+1} + By^{k+1} - b) \\ + \nabla \psi(y^{k+1}) - \nabla \psi(y^k) = 0. \end{aligned} \quad (8b)$$

And then these, by the update formula (7c), can be rewritten as

$$\nabla f(x^k) + H_k^x \Delta x^{k+1} - A^T \lambda^{k+1} - \beta A^T B \Delta y^{k+1} + \nabla \phi(x^{k+1}) - \nabla \phi(x^k) = 0, \quad (9a)$$

$$\nabla g(y^k) + H_k^y \Delta y^{k+1} - B^T \lambda^{k+1} + \nabla \psi(y^{k+1}) - \nabla \psi(y^k) = 0. \quad (9b)$$

Based on the above analysis and preparation, now we describe the proposed algorithm in detail for solving problems (1) as follows (Algorithm 1).

**Algorithm 1** (ADMM-based SQP)

**Step 0.** Input initial iterative point  $(x^0, y^0, \lambda^0)$ , select  $\beta > 0$ , initial symmetric matrices  $H_0^x \in R^{n_1 \times n_1}$  and  $H_0^y \in R^{n_2 \times n_2}$ , and continuously differentiable strongly convex functions  $\phi$  and  $\psi$ . Set  $k := 0$ .

**Step 1.** Solve QP subproblems (7a) and (7b) to obtain the (unique) optimal solution  $x^{k+1}$  and  $y^{k+1}$ , respectively, and update  $\lambda^{k+1}$  by (7c).

**Step 2.** If a termination criterion is not met, calculate new matrices  $H_{k+1}^x$  and  $H_{k+1}^y$ , set  $k := k + 1$ , and return to Step 1.

*Remark 1* At first glance, one might view the ADMM-based SQP method proposed in this paper as a special case of the well-known Bregman ADMM [10], whose iterative scheme is given as follows:

$$\begin{cases} x^{k+1} = \arg \min_{x \in R^{n_1}} \{ \mathcal{L}_\beta(x, y^k, \lambda^k) + \Delta_{\tilde{\phi}_k}(x, x^k) \}, \\ y^{k+1} = \arg \min_{y \in R^{n_2}} \{ \mathcal{L}_\beta(x^{k+1}, y, \lambda^k) + \Delta_{\tilde{\psi}_k}(y, y^k) \}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b), \end{cases}$$

when given some concrete choices of functions  $\tilde{\phi}_k(x)$  and  $\tilde{\psi}_k(y)$  such as

$$\tilde{\phi}_k(x) := \phi(x) + \frac{1}{2} \|x\|_{H_k^x}^2 - f(x), \quad \tilde{\psi}_k(y) := \psi(y) + \frac{1}{2} \|y\|_{H_k^y}^2 - g(y).$$

However, this may cause disagreements, since the pursuit of this statement raised above is simply a matter of form. In fact, the convexity property is required for a Bregman distance by definition, see, e.g., [28, 29]. Meanwhile, it is worth pointing out that matrices  $H_k^x$  and  $H_k^y$  in this paper are not required to be positive semidefinite/definite, and function  $f$  or  $g$  is also not necessarily convex.

### 3 Convergence analysis

This section is devoted to the convergence analysis of the ADMM-based SQP method introduced in Sect. 2. First, we consider some basic assumptions as follows.

**Assumption 1** Let  $\min\{\sigma_0, \sigma_\phi, \sigma_\psi\} > 0$ ,  $f: R^{n_1} \rightarrow R$ , and  $g: R^{n_2} \rightarrow R$  be continuously differentiable functions. Assume that the following conditions hold:

- (i)  $BB^T \succeq \sigma_0 I$ , namely  $B$  is full row rank;
- (ii)  $\phi$  and  $\psi$  are strongly convex with modulus  $\sigma_\phi, \sigma_\psi$ , respectively;
- (iii)  $\nabla f, \nabla g, \nabla \phi$ , and  $\nabla \psi$  are Lipschitz continuous with modulus  $\ell_f, \ell_g, \ell_\phi, \ell_\psi > 0$ , respectively.

**Assumption 2** The matrix sequences  $\{H_k^x\}$  and  $\{H_k^y\}$  are bounded, and there exist constants  $h, \eta^x, \eta^y > 0$  such that

$$\|H_k^x\| \leq h, \quad \|H_k^y\| \leq h, \quad \mathcal{H}_k^x \succeq \eta^x I_{n_1}, \quad \mathcal{H}_k^y \succeq \eta^y I_{n_2}, \quad \forall k, \quad (10)$$

where

$$\mathcal{H}_k^x := H_k^x + \sigma_\phi I_{n_1} - \ell_f I_{n_1}, \quad (11a)$$

$$\mathcal{H}_k^y := H_k^y + \sigma_\psi I_{n_2} - \left( \ell_g + \frac{4(\ell_\psi + h)^2 + 4(\ell_g + \ell_\psi + h)^2}{\beta \sigma_0} \right) I_{n_2}. \quad (11b)$$

**Remark 2** From Assumption 2, in order to guarantee that the final two relations in (10) hold, it is suitable for us to choose such quantities

$$\inf \sigma_{H_k^x} + \sigma_\phi - \ell_f \geq \eta^x,$$

and

$$\beta > \bar{\beta} := \frac{4[(\ell_\psi + h)^2 + (\ell_g + \ell_\psi + h)^2]}{\sigma_0(\inf \sigma_{H_k^y} + \sigma_\psi - \ell_g - \eta^y)} > 0, \quad \forall k, \quad (12)$$

where  $\sigma_{H_k^x}$  and  $\sigma_{H_k^y}$  are the minimum eigenvalues of matrices  $H_k^x$  and  $H_k^y$ , respectively. Besides, there is no doubt that both QP subproblems (7a) and (7b) have a unique optimal solution.

To design an appropriate merit function for problem (1), we introduce a modified *potential function*  $\hat{\mathcal{L}}_\beta : R^{n_1} \times R^{n_2} \times R^m \times R^{n_2} \rightarrow R$  defined as

$$\hat{\mathcal{L}}_\beta(\hat{w}) := \mathcal{L}_\beta(w) + \delta \|y - \hat{y}\|^2, \quad \delta := \frac{2(\ell_g + \ell_\psi + h)^2}{\beta \sigma_0}. \quad (13)$$

Before giving the descent property of  $\hat{\mathcal{L}}_\beta(\cdot)$ , we first establish a series of technical results, which shall contribute to characterizing convergence properties of the ADMM-based SQP algorithm. To see that, we now provide an upper estimate for the quantities  $\|\Delta \lambda^{k+1}\|^2$  in the following.

**Lemma 2** Suppose that Assumptions 1 and 2 are satisfied. Then we have

$$\|\Delta \lambda^{k+1}\|^2 \leq \frac{2(\ell_\psi + h)^2}{\sigma_0} \|\Delta y^{k+1}\|^2 + \frac{2(\ell_g + \ell_\psi + h)^2}{\sigma_0} \|\Delta y^k\|^2. \quad (14)$$

*Proof* First, it follows directly from Assumption 1(i) that

$$\|B^T \Delta \lambda^{k+1}\|^2 \geq \sigma_0 \|\Delta \lambda^{k+1}\|^2. \quad (15)$$

Again, by the optimality condition (9b), we have

$$B^T \lambda^{k+1} = \nabla g(y^k) + H_k^y \Delta y^{k+1} + \nabla \psi(y^{k+1}) - \nabla \psi(y^k). \quad (16)$$

Then, using inequality (16), Assumption 1(iii), and the boundedness of  $\{H_k^y\}$  in Assumption 2, we obtain

$$\begin{aligned} & \|B^T \Delta \lambda^{k+1}\|^2 \\ &= \|(\nabla g(y^k) + H_k^y \Delta y^{k+1} + \nabla \psi(y^{k+1}) - \nabla \psi(y^k)) \\ &\quad - (\nabla g(y^{k-1}) + H_{k-1}^y \Delta y^k + \nabla \psi(y^k) - \nabla \psi(y^{k-1}))\|^2 \end{aligned}$$

$$\begin{aligned}
&\leq [(\ell_\psi + h)\|\Delta y^{k+1}\| + (\ell_g + \ell_\psi + h)\|\Delta y^k\|]^2 \\
&\leq 2(\ell_\psi + h)^2\|\Delta y^{k+1}\|^2 + 2(\ell_g + \ell_\psi + h)^2\|\Delta y^k\|^2,
\end{aligned} \tag{17}$$

which together with relation (15) immediately establishes the assertion.  $\square$

To proceed, the following lemma bounds the pointwise change of the augmented Lagrangian function.

**Lemma 3** *Suppose that Assumptions 1 and 2 are satisfied, let  $\{w^k\}$  be the sequence generated by the ADMM-based SQP method. Then the following assertion is true:*

$$\begin{aligned}
&\mathcal{L}_\beta(w^{k+1}) - \mathcal{L}_\beta(w^k) \\
&\leq -\frac{1}{2}\|\Delta x^{k+1}\|_{(H_k^x + \sigma_\phi I_{n_1} - \ell_f I_{n_1})}^2 + \frac{2(\ell_g + \ell_\psi + h)^2}{\beta\sigma_0}\|\Delta y^k\|^2 \\
&\quad - \frac{1}{2}\|\Delta y^{k+1}\|_{(H_k^y + \sigma_\psi I_{n_2} - (\ell_g + \frac{4(\ell_\psi + h)^2}{\beta\sigma_0})I_{n_2})}^2.
\end{aligned} \tag{18}$$

*Proof* From Assumption 1(iii), we can see that  $\nabla f$  and  $\nabla g$  are Lipschitz continuous, then we can deduce

$$\begin{aligned}
&f(x^{k+1}) - f(x^k) - \nabla f(x^k)^T(x^{k+1} - x^k) \\
&= \int_0^1 \nabla f(x^k + s(x^{k+1} - x^k))^T(x^{k+1} - x^k) ds - \int_0^1 \nabla f(x^k)^T(x^{k+1} - x^k) ds \\
&= \int_0^1 [\nabla f(x^k + s(x^{k+1} - x^k)) - \nabla f(x^k)]^T(x^{k+1} - x^k) ds \\
&\leq \int_0^1 \|\nabla f(x^k + s(x^{k+1} - x^k)) - \nabla f(x^k)\| \cdot \|x^{k+1} - x^k\| ds \\
&\leq \ell_f \|\Delta x^{k+1}\|^2 \int_0^1 s ds = \frac{\ell_f}{2} \|\Delta x^{k+1}\|^2,
\end{aligned} \tag{19}$$

and similarly

$$g(y^{k+1}) \leq g(y^k) + \nabla g(y^k)^T(y^{k+1} - y^k) + \frac{\ell_g}{2}\|\Delta y^{k+1}\|^2. \tag{20}$$

By the definition of  $\mathcal{L}_\beta(\cdot)$  and formula (7c), we have

$$\begin{aligned}
&\mathcal{L}_\beta(x^{k+1}, y^{k+1}, \lambda^{k+1}) - \mathcal{L}_\beta(x^{k+1}, y^{k+1}, \lambda^k) \\
&= (\lambda^k - \lambda^{k+1})^T(Ax^{k+1} + By^{k+1} - b) = \frac{1}{\beta}\|\Delta \lambda^{k+1}\|^2.
\end{aligned} \tag{21}$$

Moreover, since  $y^{k+1}$  is a minimizer of (7b), then using the strong convexity of  $\psi$  and relation (20), we have

$$\mathcal{L}_\beta(x^{k+1}, y^{k+1}, \lambda^k) - \mathcal{L}_\beta(x^{k+1}, y^k, \lambda^k)$$



$$\begin{aligned}
&= f(x^{k+1}) + g(y^{k+1}) - (\lambda^k)^T (Ax^{k+1} + By^{k+1} - b) + \frac{\beta}{2} \|Ax^{k+1} + By^{k+1} - b\|^2 \\
&\quad - \left[ f(x^{k+1}) + g(y^k) - (\lambda^k)^T (Ax^{k+1} + By^k - b) + \frac{\beta}{2} \|Ax^{k+1} + By^k - b\|^2 \right] \\
&\leq g(y^{k+1}) - g(y^k) - \nabla g(y^k)^T (y^{k+1} - y^k) - \frac{1}{2} \|\Delta y^{k+1}\|_{H_k^y}^2 - \Delta \psi(y^{k+1}, y^k) \\
&\leq \frac{\ell_g}{2} \|\Delta y^{k+1}\|^2 - \frac{1}{2} \|\Delta y^{k+1}\|_{H_k^y}^2 - \frac{\sigma_\psi}{2} \|\Delta y^{k+1}\|^2, \\
&= -\frac{1}{2} \|\Delta y^{k+1}\|_{(H_k^y + \sigma_\psi I_{n_2} - \ell_g I_{n_2})}^2.
\end{aligned} \tag{22}$$

Again, similarly recalling the update for  $x$ -subproblem (7a), one also has

$$\begin{aligned}
&\mathcal{L}_\beta(x^{k+1}, y^k, \lambda^k) - \mathcal{L}_\beta(x^k, y^k, \lambda^k) \\
&= f(x^{k+1}) + g(y^k) - (\lambda^k)^T (Ax^{k+1} + By^k - b) + \frac{\beta}{2} \|Ax^{k+1} + By^k - b\|^2 \\
&\quad - \left[ f(x^k) + g(y^k) - (\lambda^k)^T (Ax^k + By^k - b) + \frac{\beta}{2} \|Ax^k + By^k - b\|^2 \right] \\
&\leq f(x^{k+1}) - f(x^k) - \nabla f(x^k)^T (x^{k+1} - x^k) - \frac{1}{2} \|\Delta x^{k+1}\|_{H_k^x}^2 - \Delta \phi(x^{k+1}, x^k) \\
&\leq \frac{\ell_f}{2} \|\Delta x^{k+1}\|^2 - \frac{1}{2} \|\Delta x^{k+1}\|_{H_k^x}^2 - \frac{\sigma_\phi}{2} \|\Delta x^{k+1}\|^2, \\
&= -\frac{1}{2} \|\Delta x^{k+1}\|_{(H_k^x + \sigma_\phi I_{n_1} - \ell_f I_{n_1})}^2.
\end{aligned} \tag{23}$$

Summing up relations (21), (22), and (23), we obtain

$$\begin{aligned}
\mathcal{L}_\beta(w^{k+1}) - \mathcal{L}_\beta(w^k) &\leq \frac{1}{\beta} \|\Delta \lambda^{k+1}\|^2 - \frac{1}{2} \|\Delta x^{k+1}\|_{(H_k^x + \sigma_\phi I_{n_1} - \ell_f I_{n_1})}^2 \\
&\quad - \frac{1}{2} \|\Delta y^{k+1}\|_{(H_k^y + \sigma_\psi I_{n_2} - \ell_g I_{n_2})}^2.
\end{aligned}$$

This, combined with relation (14), justifies the conclusion.  $\square$

We next give the descent property of  $\hat{\mathcal{L}}_\beta(\cdot)$ , i.e., the sequence  $\{\hat{\mathcal{L}}_\beta(\hat{w}^k)\}$  is monotonically nonincreasing when depending on Assumption 1 and Assumption 2.

**Lemma 4** Suppose that Assumptions 1 and 2 hold, let  $\{w^k\}$  be the sequence generated by the ADMM-based SQP method. Then we have

$$\hat{\mathcal{L}}_\beta(\hat{w}^{k+1}) \leq \hat{\mathcal{L}}_\beta(\hat{w}^k) - \frac{1}{2} \|\Delta x^{k+1}\|_{\mathcal{H}_k^x}^2 - \frac{1}{2} \|\Delta y^{k+1}\|_{\mathcal{H}_k^y}^2, \tag{24}$$

where matrices  $\mathcal{H}_k^x$  and  $\mathcal{H}_k^y$  are defined in (11a) and (11b), respectively.

*Proof* Clearly, we observe from Lemma 3 that

$$\mathcal{L}_\beta(w^{k+1}) + \frac{2(\ell_g + \ell_\psi + h)^2}{\beta \sigma_0} \|\Delta y^{k+1}\|^2$$

$$\begin{aligned} &\leq \mathcal{L}_\beta(w^k) + \frac{2(\ell_g + \ell_\psi + h)^2}{\beta\sigma_0} \|\Delta y^k\|^2 - \frac{1}{2} \|\Delta x^{k+1}\|_{(H_k^x + \sigma_\phi I_{n_1} - \ell_f I_{n_1})}^2 \\ &\quad - \frac{1}{2} \|\Delta y^{k+1}\|_{(H_k^y + \sigma_\psi I_{n_2} - (\ell_g + \frac{4(\ell_\psi + h)^2 + 4(\ell_g + \ell_\psi + h)^2}{\beta\sigma_0}) I_{n_2})}^2. \end{aligned}$$

Thus, using the definitions of  $\hat{\mathcal{L}}_\beta(\cdot)$ ,  $\mathcal{H}_k^x$ , and  $\mathcal{H}_k^y$ , the conclusion is immediately satisfied.  $\square$

Notice that the boundedness of the iterative sequence  $\{w^k\}$  generated by the ADMM-based SQP method plays an important role in the existence of a cluster point, so it is necessary to consider some additional conditions as follows.

**Assumption 3** Assume that the following conditions are satisfied:

- (i) There exists a constant  $\tilde{\sigma} > 0$  with  $g^* := \inf_y \{g(y) - \tilde{\sigma} \|\nabla g(y)\|^2\} > -\infty$ ;
- (ii)  $f$  and  $g$  are coercive, i.e.,  $\lim_{\|x\| \rightarrow +\infty} f(x) = +\infty$ ,  $\lim_{\|y\| \rightarrow +\infty} g(y) = +\infty$ ;
- (iii)  $\beta > \max\{\tilde{\beta}, \frac{1}{\tilde{\sigma}\sigma_0}\}$ .

We now prove the boundedness of the iterative sequence  $\{w^k\}$ , which is further beneficial to illustrate that the potential function constructed is bounded from below.

**Lemma 5** Suppose that Assumptions 1, 2, and 3 are all satisfied, let  $\{w^k\}$  be the sequence generated by the ADMM-based SQP method. Then  $\{w^k\}$  is bounded, and there exists a constant  $\underline{\mathcal{L}}$  such that  $\hat{\mathcal{L}}_\beta(\hat{w}^k) \geq \underline{\mathcal{L}} > -\infty, \forall k > 0$ .

*Proof* First, by matching the complete square, we have

$$\begin{aligned} \hat{\mathcal{L}}_\beta(\hat{w}^k) &= f(x^k) + g(y^k) - (\lambda^k)^T (Ax^k + By^k - b) \\ &\quad + \frac{\beta}{2} \|Ax^k + By^k - b\|^2 + \delta \|\Delta y^k\|^2 \\ &= f(x^k) + g(y^k) + \frac{\beta}{2} \left\| Ax^k + By^k - b - \frac{\lambda^k}{\beta} \right\|^2 \\ &\quad - \frac{1}{2\beta} \|\lambda^k\|^2 + \delta \|\Delta y^k\|^2. \end{aligned} \quad (25)$$

Next, recalling relation (16), Assumption 1(iii), and Assumption 2 gives

$$\begin{aligned} \sigma_0 \|\lambda^k\|^2 &\leq \|B^T \lambda^k\|^2 = \|\nabla g(y^{k-1}) + H_{k-1}^y \Delta y^k + \nabla \psi(y^k) - \nabla \psi(y^{k-1})\|^2 \\ &\leq (\|\nabla g(y^{k-1}) - \nabla g(y^k)\| + \|\nabla g(y^k)\| + (\ell_\psi + h) \|\Delta y^k\|)^2 \\ &\leq (\|\nabla g(y^k)\| + (\ell_g + \ell_\psi + h) \|\Delta y^k\|)^2 \\ &\leq 2\|\nabla g(y^k)\|^2 + 2(\ell_g + \ell_\psi + h)^2 \|\Delta y^k\|^2, \end{aligned}$$

and hence

$$\frac{1}{2\beta} \|\lambda^k\|^2 \leq \frac{1}{\beta\sigma_0} \|\nabla g(y^k)\|^2 + \frac{(\ell_g + \ell_\psi + h)^2}{\beta\sigma_0} \|\Delta y^k\|^2. \quad (26)$$

Subtracting this into (25) and invoking the definition of  $\delta$  in (13), we obtain

$$\begin{aligned}\hat{\mathcal{L}}_{\beta}(\hat{w}^k) &\geq f(x^k) + g(y^k) - \frac{1}{\beta\sigma_0} \|\nabla g(y^k)\|^2 \\ &\quad + \frac{\beta}{2} \left\| Ax^k + By^k - b - \frac{\lambda^k}{\beta} \right\|^2 + \frac{(\ell_g + \ell_{\psi} + h)^2}{\beta\sigma_0} \|\Delta y^k\|^2 \\ &= f(x^k) + (g(y^k) - \tilde{\sigma} \|\nabla g(y^k)\|^2) + \left( \tilde{\sigma} - \frac{1}{\beta\sigma_0} \right) \|\nabla g(y^k)\|^2 \\ &\quad + \frac{\beta}{2} \left\| Ax^k + By^k - b - \frac{\lambda^k}{\beta} \right\|^2 + \frac{(\ell_g + \ell_{\psi} + h)^2}{\beta\sigma_0} \|\Delta y^k\|^2.\end{aligned}\quad (27)$$

On the other hand, from Lemma 4 and Assumption 2, we know that  $\{\hat{\mathcal{L}}_{\beta}(\hat{w}^k)\}$  is nonincreasing, and thus we have

$$\hat{\mathcal{L}}_{\beta}(\hat{w}^1) \geq \hat{\mathcal{L}}_{\beta}(\hat{w}^k). \quad (28)$$

Since  $\lim_{\|x\| \rightarrow +\infty} f(x) = +\infty$  implies that  $f^* := \inf_x f(x) > -\infty$ , which together with Assumption 3 as well as relations (27) and (28), implies that sequences  $\{x^k\}$ ,  $\{\|\nabla g(y^k)\|\}$ , and  $\{\|\Delta y^k\|\}$  are bounded, and then the boundedness of  $\{\lambda^k\}$  follows directly from estimate (26). Moreover,  $\{y^k\}$  is also bounded by taking the fact that  $\lim_{\|y\| \rightarrow +\infty} g(y) = +\infty$  implies that  $\inf_y g(y) > -\infty$ . Therefore, the sequence  $\{w^k\}$  is bounded.

To this end, ignoring some nonnegative terms of (27), one has

$$\hat{\mathcal{L}}_{\beta}(\hat{w}^k) \geq f(x^k) + g(y^k) - \tilde{\sigma} \|\nabla g(y^k)\|^2 \geq \underline{\mathcal{L}},$$

where  $\underline{\mathcal{L}} := f^* + g^*$ , and the proof is completed.  $\square$

Now, we are ready to establish the global convergence of the ADMM-based SQP method.

**Theorem 1** *Suppose that Assumptions 1, 2, and 3 are all satisfied, let  $\{w^k\}$  be the sequence generated by the ADMM-based SQP method. Then*

- (i)  $\lim_{k \rightarrow \infty} (\|\Delta x^{k+1}\| + \|\Delta y^{k+1}\| + \|\Delta \lambda^{k+1}\|) = 0$ ;
- (ii) *any cluster point  $w^*$  of the sequence  $\{w^k\}$  is a KKT point of problem (1).*

*Proof* (i) Note first from Lemma 4 that

$$\hat{\mathcal{L}}_{\beta}(\hat{w}^k) - \hat{\mathcal{L}}_{\beta}(\hat{w}^{k+1}) \geq \frac{1}{2} \|\Delta x^{k+1}\|_{\mathcal{H}_k^x}^2 + \frac{1}{2} \|\Delta y^{k+1}\|_{\mathcal{H}_k^y}^2.$$

Thus, summing up this inequality from  $k = 1$  to  $n$ , using Lemma 5 yields

$$\begin{aligned}\infty &> \hat{\mathcal{L}}_{\beta}(\hat{w}^1) - \underline{\mathcal{L}} \geq \hat{\mathcal{L}}_{\beta}(\hat{w}^1) - \hat{\mathcal{L}}_{\beta}(\hat{w}^{n+1}) \\ &\geq \frac{1}{2} \sum_{k=1}^n \|\Delta x^{k+1}\|_{\mathcal{H}_k^x}^2 + \frac{1}{2} \sum_{k=1}^n \|\Delta y^{k+1}\|_{\mathcal{H}_k^y}^2.\end{aligned}\quad (29)$$

Again, by Assumption 2, we know that, for any  $k$ , both matrices  $\mathcal{H}_k^x$  and  $\mathcal{H}_k^y$  are positive definite. Thus

$$\sum_{k=1}^{\infty} \|\Delta x^{k+1}\|^2 < \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \|\Delta y^{k+1}\|^2 < \infty.$$

Furthermore, in view of estimate (14), one also has  $\sum_{k=1}^{\infty} \|\Delta \lambda^{k+1}\|^2 < \infty$ . So,

$$\|\Delta x^{k+1}\| \rightarrow 0, \quad \|\Delta y^{k+1}\| \rightarrow 0, \quad \|\Delta \lambda^{k+1}\| \rightarrow 0, \quad k \rightarrow \infty, \quad (30)$$

which completes assertion (i).

(ii) By Lemma 5, we argue that  $\{w^k\}$  is bounded and thus there exists at least one cluster point. Assume that  $w^*$  is a cluster point of  $\{w^k\}$ , and let  $\{w^{k_j}\}$  be a convergent subsequence such that  $\lim_{j \rightarrow \infty} w^{k_j} = w^*$ . On the other hand, from Assumption 1(iii) and assertion (i), we know

$$\nabla \phi(x^{k+1}) - \nabla \phi(x^k) \rightarrow 0, \quad \nabla \psi(y^{k+1}) - \nabla \psi(y^k) \rightarrow 0, \quad k \rightarrow \infty. \quad (31)$$

In view of this, and taking limit in (7c), (9a), and (9b) along the sequence  $\{w^{k_j}\}$ , we obtain

$$\nabla f(x^*) - A^T \lambda^* = 0, \quad \nabla g(y^*) - B^T \lambda^* = 0, \quad Ax^* + By^* = b.$$

This implies that  $w^*$  is a KKT point of problem (1).  $\square$

It is well known that when the potential function has a geometric property known as the KL property, then Theorem 1 can typically be strengthened since the limit point of the iterative sequence is unique.

**Theorem 2** *Suppose that Assumptions 1, 2, and 3 hold, and suppose that  $f$  and  $g$  are semi-algebraic functions, let  $\{w^k\}$  be the sequence generated by the ADMM-based SQP method. Then the whole sequence  $\{w^k\}$  converges to a KKT point of problem (1).*

*Proof* Clearly, by Theorem 1, it suffices to prove that the sequence  $\{w^k\}$  is convergent. From Lemmas 4 and 5, we know that  $\{\hat{w}^k\}$  is also bounded and  $\{\hat{\mathcal{L}}_\beta(\hat{w}^k)\}$  is monotonically bounded from below, so  $\mathcal{L}^* := \lim_{k \rightarrow \infty} \hat{\mathcal{L}}_\beta(\hat{w}^k)$  should exist. Then the rest of the proof is divided into two cases.

Suppose first that  $\hat{\mathcal{L}}_\beta(\hat{w}^N) = \mathcal{L}^*$  for some  $N \geq 1$ . Since  $\{\hat{\mathcal{L}}_\beta(\hat{w}^k)\}$  is nonincreasing, we have  $\hat{\mathcal{L}}_\beta(\hat{w}^k) = \mathcal{L}^*$  for any  $k \geq N$ . Then, according to Lemma 4,  $x^{N+t} = x^N$  and  $y^{N+t} = y^N$  hold for any  $t \geq 0$ . This implies that both sequences  $\{x^k\}$  and  $\{y^k\}$  converge finitely. Furthermore, it follows from relation (14) that  $\lambda^{N+t} = \lambda^{N+1}$  for any  $t \geq 1$ . Hence,  $\{\lambda^k\}$  is also convergent, which justifies the assertion.

Suppose next that  $\hat{\mathcal{L}}_\beta(\hat{w}^k) > \mathcal{L}^*$  for all  $k$ . Note that  $\hat{\mathcal{L}}_\beta(\cdot)$  is a semialgebraic function due to the semialgebraicity of  $f$  and  $g$ , hence it is a KL function. Subsequently, this part involves three steps for analysis.

To begin with, we prove that  $\hat{\mathcal{L}}_\beta(\cdot)$  is constant on  $\Omega$ , where  $\Omega$  is a set of all cluster points of the sequence  $\{\hat{w}^k\}$ , and then utilize the uniformized KL property in Lemma 1.

By the boundedness of  $\{\hat{w}^k\}$ , obviously,  $\Omega$  is nonempty and  $d(\hat{w}^k, \Omega) \rightarrow 0$ . Similarly, as proved in [24, Lemma 5(iii)], we also know that  $\Omega$  is a compact and connected set. Let  $\hat{w}^* \in \Omega$  be arbitrary, and consider a convergent subsequence  $\{\hat{w}^{k_j}\}$  of  $\{\hat{w}^k\}$  converging to  $\hat{w}^*$ . Then, by the continuity of  $\hat{\mathcal{L}}_\beta(\cdot)$ , we have  $\mathcal{L}^* = \lim_{j \rightarrow \infty} \hat{\mathcal{L}}_\beta(\hat{w}^{k_j}) = \hat{\mathcal{L}}_\beta(\hat{w}^*)$ , so  $\{\hat{\mathcal{L}}_\beta(\hat{w}^k)\}$  is convergent, namely  $\hat{\mathcal{L}}_\beta(\hat{w}^k) \rightarrow \hat{\mathcal{L}}_\beta(\hat{w}^*)$ . Thus,  $\hat{\mathcal{L}}_\beta(\cdot)$  is constant on  $\Omega$  since  $\hat{w}^* \in \Omega$  is arbitrary. Now, using Lemma 1, for any  $\delta_0 > 0$ ,  $\eta > 0$ , there exists an integer  $k_1 > 0$  such that  $d(\hat{w}^k, \Omega) < \delta_0$  and  $\mathcal{L}^* < \hat{\mathcal{L}}_\beta(\hat{w}^k) < \mathcal{L}^* + \eta$ , and

$$\varphi'(\hat{\mathcal{L}}_\beta(\hat{w}^k) - \mathcal{L}^*)d(0, \nabla \hat{\mathcal{L}}_\beta(\hat{w}^k)) \geq 1, \quad \forall k \geq k_1. \quad (32)$$

Next, we attempt to bound the distance from 0 to  $\nabla \hat{\mathcal{L}}_\beta(\hat{w}^k)$ . Taking the partial derivative of  $\hat{\mathcal{L}}_\beta(\cdot)$  at  $\hat{w}^k$  with respect to variable  $x$ , we have

$$\begin{aligned} \nabla_x \hat{\mathcal{L}}_\beta(\hat{w}^k) &= \nabla f(x^k) - A^T \lambda^k + \beta A^T (Ax^k + By^k - b) \\ &= -H_k^x \Delta x^{k+1} - \beta A^T A \Delta x^{k+1} - (\nabla \phi(x^{k+1}) - \nabla \phi(x^k)), \end{aligned} \quad (33)$$

where the final equality follows from the optimality condition (8a). Likewise, for variable  $y$ , we have

$$\begin{aligned} \nabla_y \hat{\mathcal{L}}_\beta(\hat{w}^k) &= \nabla g(y^k) - B^T \lambda^k + \beta B^T (Ax^k + By^k - b) + 2\delta \Delta y^k \\ &= -H_k^y \Delta y^{k+1} - (\nabla \psi(y^{k+1}) - \nabla \psi(y^k)) \\ &\quad - \beta B^T (A \Delta x^{k+1} + B \Delta y^{k+1}) + 2\delta \Delta y^k \\ &= -H_k^y \Delta y^{k+1} - (\nabla \psi(y^{k+1}) - \nabla \psi(y^k)) \\ &\quad + B^T (\Delta \lambda^{k+1} - \Delta \lambda^k) + 2\delta \Delta y^k, \end{aligned} \quad (34)$$

where the second equality follows from (8b), and the final equality utilizes formula (7c). Additionally,

$$\nabla_\lambda \hat{\mathcal{L}}_\beta(\hat{w}^k) = -(Ax^k + By^k - b) = \frac{1}{\beta} \Delta \lambda^k \quad (35)$$

and

$$\nabla_{\hat{y}} \hat{\mathcal{L}}_\beta(\hat{w}^k) = -2\delta \Delta y^k. \quad (36)$$

Since  $\nabla \phi$  and  $\nabla \psi$  are Lipschitz continuous, then combining Assumption 2 with relations (33)–(36), there exists a constant  $a > 0$  such that

$$d(0, \nabla \hat{\mathcal{L}}_\beta(\hat{w}^k)) \leq a(\|\Delta x^{k+1}\| + \|\Delta y^{k+1}\| + \|\Delta y^k\| + \|\Delta \lambda^{k+1}\| + \|\Delta \lambda^k\|). \quad (37)$$

Finally, based on relations (32) and (37), we start to study the convergence of the entire sequence  $\{w^k\}$ . For statement convenience, let

$$\Delta^k := \varphi(\hat{\mathcal{L}}_\beta(\hat{w}^k) - \mathcal{L}^*) - \varphi(\hat{\mathcal{L}}_\beta(\hat{w}^{k+1}) - \mathcal{L}^*)$$

and

$$\ell^k := \|\Delta x^{k+1}\| + \|\Delta y^{k+1}\| + \|\Delta y^k\| + \|\Delta \lambda^{k+1}\| + \|\Delta \lambda^k\|.$$

As  $\hat{\mathcal{L}}_\beta(\cdot)$  is nonincreasing and by Definition 1 that  $\varphi$  is monotonous, it is easy to have  $\Delta^k \geq 0$  for  $k \geq k_1$ . Multiplying both sides of (37) by  $\Delta^k$ , using inequality (32) and the concavity of  $\varphi$ , as well as Lemma 4 and Assumption 2, we obtain for all  $k \geq k_1$  that

$$\begin{aligned} \ell^k \cdot (a\Delta^k) &\geq d(0, \nabla \hat{\mathcal{L}}_\beta(\hat{w}^k)) \cdot \Delta^k \\ &\geq d(0, \nabla \hat{\mathcal{L}}_\beta(\hat{w}^k)) \cdot \varphi'(\hat{\mathcal{L}}_\beta(\hat{w}^k) - \mathcal{L}^*) \cdot (\hat{\mathcal{L}}_\beta(\hat{w}^k) - \hat{\mathcal{L}}_\beta(\hat{w}^{k+1})) \\ &\geq \hat{\mathcal{L}}_\beta(\hat{w}^k) - \hat{\mathcal{L}}_\beta(\hat{w}^{k+1}) \geq \frac{1}{2}(\|\Delta x^{k+1}\|_{\mathcal{H}_k^x}^2 + \|\Delta y^{k+1}\|_{\mathcal{H}_k^y}^2) \\ &\geq \frac{\eta^x}{2}\|\Delta x^{k+1}\|^2 + \frac{\eta^y}{2}\|\Delta y^{k+1}\|^2 \\ &\geq \frac{1}{2}\min\{\eta^x, \eta^y\}(\|\Delta x^{k+1}\|^2 + \|\Delta y^{k+1}\|^2) \\ &\geq \frac{1}{4}\min\{\eta^x, \eta^y\}(\|\Delta x^{k+1}\| + \|\Delta y^{k+1}\|)^2 \\ &:= \sigma_1(\|\Delta x^{k+1}\| + \|\Delta y^{k+1}\|)^2, \end{aligned} \quad (38)$$

where  $\sigma_1 := \frac{1}{4}\min\{\eta^x, \eta^y\} > 0$ . Now, dividing both sides of (38) by  $\sigma_1$ , taking the square root, and then applying the inequality  $\frac{u+v}{2} \geq \sqrt{uv}$  for  $u, v \geq 0$  to address the resulting inequality, we get

$$\frac{\ell^k}{2m} + \frac{am}{2\sigma_1}\Delta^k \geq \|\Delta x^{k+1}\| + \|\Delta y^{k+1}\|, \quad k \geq k_1, \quad (39)$$

where  $m > 0$  is an arbitrary constant. Then, by relation (16), Assumptions 1 and 2, we can derive

$$\begin{aligned} \sqrt{\sigma_0}\|\Delta \lambda^{k+1}\| &\leq \|B^T \Delta \lambda^{k+1}\| \\ &= \|(\nabla g(y^k) + H_k^y \Delta y^{k+1} + \nabla \psi(y^{k+1}) - \nabla \psi(y^k)) \\ &\quad - (\nabla g(y^{k-1}) + H_{k-1}^y \Delta y^k + \nabla \psi(y^k) - \nabla \psi(y^{k-1}))\| \\ &\leq (\ell_g + \ell_\psi + h)\|\Delta y^k\| + (\ell_\psi + h)\|\Delta y^{k+1}\|. \end{aligned}$$

Accordingly, for simplicity, let  $d_1 := \frac{\ell_g + \ell_\psi + h}{\sqrt{\sigma_0}}$  and  $d_2 := \frac{\ell_\psi + h}{\sqrt{\sigma_0}}$ , we have

$$\|\Delta \lambda^{k+1}\| \leq d_1 \|\Delta y^k\| + d_2 \|\Delta y^{k+1}\|, \quad (40)$$

and

$$\|\Delta \lambda^k\| \leq d_1 \|\Delta y^{k-1}\| + d_2 \|\Delta y^k\|.$$

Hence, substituting these into (39) and regrouping terms, we obtain

$$\left(1 - \frac{1}{2m}\right)\|\Delta x^{k+1}\| + \left(1 - \frac{1+d_2}{2m}\right)\|\Delta y^{k+1}\|$$

$$\leq \frac{1+d_1+d_2}{2m} \|\Delta y^k\| + \frac{d_1}{2m} \|\Delta y^{k-1}\| + \frac{am}{2\sigma_1} \Delta^k,$$

which by manipulating, can be rewritten as

$$\begin{aligned} & \left(1 - \frac{1}{2m}\right) \|\Delta x^{k+1}\| + \left(1 - \frac{1+d_1+d_2}{m}\right) \|\Delta y^{k+1}\| \\ & \leq \frac{1+2d_1+d_2}{2m} (\|\Delta y^k\| - \|\Delta y^{k+1}\|) + \frac{d_1}{2m} (\|\Delta y^{k-1}\| - \|\Delta y^k\|) + \frac{am}{2\sigma_1} \Delta^k. \end{aligned} \quad (41)$$

As a result, making use of the nonnegativity of  $\varphi$ , summing up inequality (41) from  $k = k_1$  to  $\infty$ , we have

$$\begin{aligned} & \left(1 - \frac{1}{2m}\right) \sum_{k=k_1}^{\infty} \|\Delta x^{k+1}\| + \left(1 - \frac{1+d_1+d_2}{m}\right) \sum_{k=k_1}^{\infty} \|\Delta y^{k+1}\| \\ & \leq \frac{1+2d_1+d_2}{2m} \|\Delta y^{k_1}\| + \frac{d_1}{2m} \|\Delta y^{k_1-1}\| + \frac{am}{2\sigma_1} \varphi(\hat{\mathcal{L}}_{\beta}(\hat{w}^{k_1}) - \mathcal{L}^*). \end{aligned} \quad (42)$$

Note that  $m > 0$  is arbitrary, now selecting  $m > 1 + d_1 + d_2$ , and hence  $1 - \frac{1}{2m} > 0$  and  $1 - \frac{1+d_1+d_2}{m} > 0$ , then it follows directly from (42) that

$$\sum_{k=k_1}^{\infty} \|\Delta x^{k+1}\| < \infty \quad \text{and} \quad \sum_{k=k_1}^{\infty} \|\Delta y^{k+1}\| < \infty.$$

That is, sequences  $\{x^k\}$  and  $\{y^k\}$  are convergent. Moreover, summing up relation (40) from  $k = k_1$  to  $\infty$ , we also receive

$$\sum_{k=k_1}^{\infty} \|\Delta \lambda^{k+1}\| \leq d_1 \sum_{k=k_1}^{\infty} \|\Delta y^k\| + d_2 \sum_{k=k_1}^{\infty} \|\Delta y^{k+1}\| < \infty,$$

which implies that  $\{\lambda^k\}$  is convergent too. That is,  $\{w^k\}$  is convergent. Therefore, combining with Theorem 1, the whole proof is finished.  $\square$

#### 4 Conclusions

In this paper, an ADMM-based SQP method for separably smooth nonconvex problems with linear equality constraints is proposed. Incorporating the favorable ideas of SQP method and the classical ADMM, the QP subproblem of the original problem is split into smaller-scale QP subproblems, which can be easily solved with the help of Bregman distances, and hence relieve the difficulty brought by solving large-scale QP itself corresponding to the primal nonconvex optimization problem. Additionally, we update the Lagrangian multipliers in a dual ascent step. Based on the KL property and other standard assumptions, the proposed method is globally and strongly convergent in terms of the potential function.

As future work, it is tempting for us to consider whether these theoretical results can be used to develop a relaxation factor in the multipliers updating (7c) or some accelerated technique for variants of the ADMM-based SQP method. This is an interesting issue worthy of further investigation.

### Acknowledgements

The authors are very grateful to the referees for their valuable comments and suggestions on the early version of the paper.

### Funding

This work was supported by the National Natural Science Foundation of China (No. 11771383), the Natural Science Foundation of Guangxi Province (No. 2018GXNSFFA281007), and the Middle-aged and Young Teachers' Basic Ability Promotion Project of Guangxi Province (No. 2017KY0537).

### Availability of data and materials

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

ML conceived of the description of the ADMM-based SQP method, the convergence analysis and drafted the manuscript. JJ carried out the idea of this paper and participated in the convergence analysis. All authors read and approved the final manuscript.

### Author details

<sup>1</sup>Department of Mathematics, Shanghai University, Shanghai, P.R. China. <sup>2</sup>College of Science, Guangxi University for Nationalities, Nanning, P.R. China.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 13 January 2020 Accepted: 18 March 2020 Published online: 30 March 2020

### References

1. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**(1), 1–122 (2011)
2. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximations. *Comput. Math. Appl.* **2**, 17–40 (1976)
3. Glowinski, R., Marocco, A.: Sur l'approximation par éléments finis d'ordre un et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires. *RAIRO. Anal. Numér.* **9**(2), 41–76 (1975)
4. Goldfarb, D., Ma, S.Q., Scheinberg, K.: Fast alternating linearization methods for minimizing the sum of two convex functions. *Math. Program.* **141**(1–2), 349–382 (2013)
5. Goldstein, T., Donoghue, B.O., Setzer, S.: Fast alternating direction optimization methods. *SIAM J. Imaging Sci.* **7**(3), 1588–1623 (2014)
6. He, B.S., Liu, H., Wang, Z.R., Yuan, X.M.: A strictly contractive Peaceman–Rachford splitting method for convex programming. *SIAM J. Optim.* **24**, 1011–1040 (2014)
7. He, B.S., Yuan, X.M.: On the  $\mathcal{O}(\frac{1}{n})$  convergence rate of the Douglas–Rachford alternating direction method. *SIAM J. Numer. Anal.* **50**(2), 700–709 (2012)
8. He, B.S., Yuan, X.M.: On non-ergodic convergence rate of Douglas–Rachford alternating direction method of multipliers. *Numer. Math.* **130**(3), 567–577 (2015)
9. Monteiro, R.D.C., Svaiter, B.F.: Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. *SIAM J. Optim.* **23**(1), 475–507 (2013)
10. Wang, F.H., Xu, Z.B., Xu, H.K.: Convergence of alternating direction method with multipliers for nonconvex composite problems (2014) Preprint. Available at: [arXiv:1410.8625](https://arxiv.org/abs/1410.8625)
11. Li, G.Y., Pong, T.K.: Global convergence of splitting methods for nonconvex composition optimization. *SIAM J. Optim.* **25**(1), 2434–2460 (2015)
12. Hong, M.Y., Luo, Z.Q., Razaviyayn, M.: Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. *SIAM J. Optim.* **26**(1), 337–364 (2016)
13. Wilson, R.B.: A simplicial method for concave programming. PhD thesis, Graduate School of Business Administration, Harvard University, Cambridge (1963)
14. Robinson, S.: A quadratically convergent algorithm for general nonlinear programming problems. *Math. Program.* **3**, 145–156 (1972)
15. Fukushima, M.: A successive quadratic programming algorithm with global and superlinear convergence properties. *Math. Program.* **35**, 253–264 (1986)
16. Solodov, M.V.: Global convergence of an SQP method without boundedness assumptions on any of the iterative sequences. *Math. Program., Ser. A* **118**, 1–12 (2009)
17. Panier, E.R., Tits, A.L.: A superlinearly convergent feasible method for the solution of inequality constrained optimization problems. *SIAM J. Control Optim.* **25**, 934–950 (1987)
18. Zhu, Z.B., Jian, J.B.: An efficient feasible SQP algorithm for inequality constrained optimization. *Nonlinear Anal., Real World Appl.* **10**, 1220–1228 (2009)
19. Jian, J.B.: A superlinearly convergent implicit smooth SQP algorithm for mathematical programs with nonlinear complementarity constraints. *Comput. Optim. Appl.* **31**(3), 335–361 (2005)
20. Jian, J.B., Zheng, H.Y., Tang, C.M., Hu, Q.J.: A new superlinearly convergent norm-relaxed method of strongly sub-feasible direction for inequality constrained optimization. *Appl. Math. Comput.* **182**, 955–976 (2006)



21. Jian, J.B.: Fast Algorithms for Smooth Constrained Optimization—Theoretical Analysis and Numerical Experiments. Science Press, Beijing (2010)
22. Jian, J.B., Lao, Y.X., Chao, M.T., Ma, G.D.: ADMM-SQP algorithm for two blocks linear constrained nonconvex optimization. *Oper. Res. Trans.* **22**(2), 79–92 (2018)
23. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka–Łojasiewicz inequality. *Math. Oper. Res.* **35**, 438–457 (2010)
24. Attouch, H., Bolte, J., Svaiter, B.F.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Program.* **146**(1–2), 459–494 (2014)
25. Bolte, J., Daniilidis, A., Lewis, A.: The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM J. Optim.* **17**, 1205–1223 (2006)
26. Bolte, J., Daniilidis, A., Lewis, A., Shiota, M.: Clarke subgradients of stratifiable functions. *SIAM J. Optim.* **18**, 556–572 (2007)
27. Attouch, H., Bolte, J., Svaiter, B.F.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss–Seidel methods. *Math. Program.* **137**, 91–129 (2013)
28. Bregman, L.: The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math. Math. Phys.* **7**, 200–217 (1967)
29. Chen, G., Teboulle, M.: Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM J. Optim.* **3**(3), 538–543 (1993)

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)