A new LQP alternating direction method for solving variational inequality problems with separable structure

Abdellah Bnouhachem[a,1] and **Abdelouahed Hamdi**[c]

[a]Ibn Zohr University, ENSA, BP 1136, Agadir, Morocco.

[c]Department of Mathematics, Statistics and Physics College of Arts and Sciences

Qatar University,P.O. Box 2713, Doha, Qatar.

Abstract. We presented a new logarithmic-quadratic proximal alternating direction scheme for the separable constrained convex programming problem. The predictor is obtained by solving series of related systems of nonlinear equations in a parallel wise. The new iterate is obtained by searching the optimal step size along a new descent direction. The new direction is obtained by the linear combination of two descent directions. Global convergence of the proposed method is proved under certain assumptions. We also reported some numerical results to illustrate the efficiency of the proposed method.

Key word. Variational inequalities; Monotone operator; logarithmic-quadratic proximal method; projection method; alternating direction method.

2000 AMS Subject Classification: 49J40, 65N30.

1 Introduction

This paper considers the constrained convex programming problem with the following separate structure:

$$\min \left\{ \theta_1(x) + \theta_2(y) | Ax + By = b, x \in \mathcal{R}^n_+, y \in \mathcal{R}^m_+ \right\}, \tag{1.1}$$

where $\theta_1 : \mathcal{R}^n_+ \to \mathcal{R}$ and $\theta_2 : \mathcal{R}^m_+ \to \mathcal{R}$ are closed proper convex functions not necessarily smooth, $A \in \mathcal{R}^{l\times n}$, $B \in \mathcal{R}^{l\times m}$ are given matrices and $b \in \mathcal{R}^l$ is a given vector.

A very rich class of applications may be modeled as problem (1.1). In practice these classes

[1]Corresponding author. Fax: +212 528232007.

E-mail address: babedallah@yahoo.com (A. Bnouhachem).

of problems have very large size and due to their practical importance, they have received a great deal of attention from many researchers. Various methods have been suggested, a popular approach is the alternating direction method (ADM) which was proposed by Gabay and Mercier [14] and Gabay [13]. The ADM can reduce the scale of variational inequalities by decomposing the original problem into a series of subproblems with a lower scale. To make the ADM more efficient and practical some strategies have been studied, for more details, one can refer [7, 10, 18, 21, 22, 26, 29].

Let $\partial(.)$ denote the sub-gradient operator of a convex function, and $f(x) \in \partial\theta_1(x)$ and $g(y) \in \partial\theta_2(y)$ are the sub-gradient of $\theta_1(x)$ and $\theta_2(y)$, respectively. By attaching a Lagrange multiplier vector $\lambda \in \mathcal{R}^l$ to the linear constraint $Ax + By = b$, problem (1.1) can be written in terms of finding $w \in \mathcal{W}$ such that

$$(w' - w)^\top Q(w) \geq 0, \quad \forall\, w' \in \mathcal{W}, \tag{1.2}$$

where

$$w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix} \quad Q(w) = \begin{pmatrix} f(x) - A^\top \lambda \\ g(y) - B^\top \lambda \\ Ax + By - b \end{pmatrix}, \quad \mathcal{W} = \mathcal{R}^n_+ \times \mathcal{R}^m_+ \times \mathcal{R}^l. \tag{1.3}$$

Problem (1.2)–(1.3) is referred to as *structured variational inequalities* (in short, SVI).
Very recently, Yuan and Li [30] developed the following logarithmic-quadratic proximal (LQP)-based decomposition method by applying the LQP terms to regularize the ADM subproblem-s: For a given $w^k = (x^k, y^k, \lambda^k) \in \mathcal{R}^n_{++} \times \mathcal{R}^m_{++} \times \mathcal{R}^l$, and $\mu \in (0, 1)$, the new iterative $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ is obtained via solving the following system:

$$f(x) - A^\top \Big[\lambda^k - H(Ax + By^k - b)\Big] + R\Big[(x - x^k) + \mu(x^k - X_k^2 x^{-1})\Big] = 0, \tag{1.4}$$

$$g(y) - B^\top \Big[\lambda^k - H(Ax + By - b)\Big] + S\Big[(y - y^k) + \mu(y^k - Y_k^2 y^{-1})\Big] = 0, \tag{1.5}$$

$$\lambda^{k+1} = \lambda^k - H(Ax^k + By^k - b), \tag{1.6}$$

where $H \in \mathcal{R}^{l\times l}$, $R \in \mathcal{R}^{n\times n}$, and $S \in \mathcal{R}^{m\times m}$ are symmetric positive definite.
Note that the LQP method was presented originally in [1]. It seems that it is easier to solve a series of systems of nonlinear equations than to solve a series of sub-variational inequalities in many cases. Later, Bnouhachem et al. [3, 4] and Li [24] proposed some LQP alternating direction methods and made the LQP alternating direction method more practical. Each iteration of the above methods contains a prediction and a correction, the predictor is obtained via solving (1.4)-(1.6) and the new iterate is obtained by a convex combination of the previous

point and the one generated by a projection-type method along a descent direction for [3, 24], while the new iterate is computed directly by an explicit formula derived from the original LQP method for [4]. The main disadvantage of the methods in [2, 3, 4, 5, 24, 30] is that solving the equation (1.5) requires the solution of equation (1.4). Hence, the alternating direction methods are not eligible for parallel computing in the sense that the solutions of (1.4)-(1.5) cannot be obtained simultaneously. This characteristic excludes the possibility of applying some advanced computing technologies to solve (1.4)-(1.5).

To overcome this difficulty, we propose a parallel descent LQP alternating direction method for solving SVI. The main advantage of the proposed method is that the predictor is obtained via solving a system of nonlinear equations in a parallel wise and the new iterate is obtained by searching the optimal step size along the integrated descent direction from two descent directions. Our results can be viewed as significant extensions of the previously known results.

2 The proposed method

This section states some preliminaries that are useful later. The first lemma provides some basic properties of projection onto Ω.

Lemma 2.1 *Let G be a symmetry positive definite matrix and Ω be a nonempty closed convex subset of R^l, we denote by $P_{\Omega,G}(.)$ the projection under the G-norm, that is,*

$$P_{\Omega,G}(v) = argmin\{\|v-u\|_G \ : \ u \in \Omega\}.$$

Then, we have the following inequalities.

$$(z - P_{\Omega,G}[z])^\top G(P_{\Omega,G}[z] - v) \geq 0, \quad \forall\, z \in R^l, v \in \Omega; \tag{2.1}$$

$$\|P_{\Omega,G}[u] - P_{\Omega,G}[v]\|_G \leq \|u-v\|_G, \quad \forall\, u, v \in R^l; \tag{2.2}$$

$$\|u - P_{\Omega,G}[z]\|_G^2 \leq \|z-u\|_G^2 - \|z - P_{\Omega,G}[z]\|_G^2, \quad \forall\, z \in R^l, u \in \Omega. \tag{2.3}$$

We make the following standard assumptions.

Assumption A. f is monotone with respect to $\mathcal{R}_+^n$ and g is monotone with respect to $\mathcal{R}_+^m$,

Assumption B. The solution set of SVI, denoted by $\mathcal{W}^*$, is nonempty.

We propose the following parallel LQP alternating direction method for solving SVI:

Algorithm 2.1.

Step 0. *The initial step:*

Given $\varepsilon > 0$, $\mu \in (0,1), \beta_1 \geq 0, \beta_2 \geq 0$ $(\beta_1 + \beta_2 > 0)$ and $w^0 = (x^0, y^0, \lambda^0) \in \mathcal{R}^n_{++} \times \mathcal{R}^m_{++} \times \mathcal{R}^l$. Set $k = 0$.

Step 1. *Prediction step:*

Compute $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \mathcal{R}^n_{++} \times \mathcal{R}^m_{++} \times \mathcal{R}^l$ by solving the following system:

$$f(x) - A^T[\lambda^k - H(Ax + By^k - b)] + R[(x - x^k) + \mu(x^k - X_k^2 x^{-1})] = 0, \tag{2.4}$$

$$g(y) - B^T[\lambda^k - H(Ax^k + By - b)] + S[(y - y^k) + \mu(y^k - Y_k^2 y^{-1})] = 0, \tag{2.5}$$

$$\tilde{\lambda}^k = \lambda^k - H(A\tilde{x}^k + B\tilde{y}^k - b). \tag{2.6}$$

Step 2. *Convergence verification:*

If $\max\{\|x^k - \tilde{x}^k\|_\infty, \|y^k - \tilde{y}^k\|_\infty, \|\lambda^k - \tilde{\lambda}^k\|_\infty\} < \epsilon$, then stop.

Step 3. *Correction step:*

The new iterate $w^{k+1}(\alpha_k) = (x^{k+1}, y^{k+1}, \lambda^{k+1})$ is given by:

$$w^{k+1}(\alpha_k) = (1 - \sigma)w^k + \sigma P_{\mathcal{W}}[w^k - \alpha_k G^{-1} d(w^k, \tilde{w}^k)], \quad \sigma \in (0,1) \tag{2.7}$$

where

$$\alpha_k = \frac{\varphi_k}{(\beta_1 + \beta_2)\|w^k - \tilde{w}^k\|_G^2}, \tag{2.8}$$

$$\varphi_k = \|w^k - \tilde{w}^k\|_M^2 + (\lambda^k - \tilde{\lambda}^k)^T(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)), \tag{2.9}$$

$$d(w^k, \tilde{w}^k) = \beta_1 D(w^k, \tilde{w}^k) + \beta_2 G(w^k - \tilde{w}^k),$$

$$D(w^k, \tilde{w}^k) = \begin{pmatrix} f(\tilde{x}^k) - A^T\tilde{\lambda}^k + A^T H(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)) \\ g(\tilde{y}^k) - B^T\tilde{\lambda}^k + B^T H(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)) \\ A\tilde{x}^k + B\tilde{y}^k - b \end{pmatrix}$$

and

$$G = \begin{pmatrix} (1+\mu)R + A^T HA & 0 & 0 \\ 0 & (1+\mu)S + B^T HB & 0 \\ 0 & 0 & H^{-1} \end{pmatrix}, M = \begin{pmatrix} R + A^T HA & 0 & 0 \\ 0 & S + B^T HB & 0 \\ 0 & 0 & H^{-1} \end{pmatrix}.$$

Set $k := k + 1$ and go to Step1.

Remark 2.1 *The main disadvantage of the methods proposed in [2, 3, 4, 5, 24, 30] is that the unknown vectors x and y in (1.5) are overlapped. Note that the eligibility of parallel computing is particularly preferable when the involved LQP system in (1.4)-(1.5) are large scale and thus time consuming. So, we propose to solve systems of nonlinear equations (2.4)-(2.5) in a parallel wise.*

By using as special case of our method, we can obtain some new LQP alternating methods for example:

- *If $x^{k+1} = \tilde{x}^k$, $y^{k+1} = \tilde{y}^k$ and $\lambda^{k+1} = \tilde{\lambda}^k$ in (2.4), (2.5) and (2.6), respectively, we obtain a new method which is different from that proposed in [30]. In our proposed method, problems (2.4) and (2.5), which produce $\tilde{x}^k$ and $\tilde{y}^k$, are parallel decomposed.*

- *If $\beta_1 = 0$ and $\beta_2 = 1$, we obtain a new method, the new iterate is obtained along a new descent direction $(w^k - \tilde{w}^k)$. Also the new iterate in [24] is obtained along the descent direction $(w^k - \tilde{w}^k)$. But two descent directions are different, problems (2.4) and (2.5), which produce the first descent direction $(w^k - \tilde{w}^k)$, are parallel decomposed. While the vectors $\tilde{x}^k$ and $\tilde{y}^k$ in problems (1.4) and (1.5), which offer the second descent direction $(w^k - \tilde{w}^k)$, are overlapped.*

- *If $\beta_1 = 1$ and $\beta_2 = 0$, we obtain the method proposed in [6].*

Therefore, the new algorithm is expected to be widely applicable.

Remark 2.2 *It is easy to check that $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ is solution of SVI if and only if*

$$\begin{cases} x^k - \tilde{x}^k = 0, \\ y^k - \tilde{y}^k = 0, \\ \lambda^k - \tilde{\lambda}^k = 0. \end{cases}$$

Hence, the stopping criterion adopted here is reasonable: if it is satisfied with a small ϵ, we can regard the current iterate as an approximate solution.

Remark 2.3 *We use the convex combination of w^k and $P_{\mathcal{W}}[w^k - \alpha_k G^{-1} d(w^k, \tilde{w}^k)]$ in (2.7) to ensure that the elements (x^{k+1}, y^{k+1}) of the new iterate $w^{k+1}(\alpha_k) = (x^{k+1}, y^{k+1}, \lambda^{k+1})$ lie in $\in \mathcal{R}^n_{++} \times \mathcal{R}^m_{++}$.*

We need the following result in the convergence analysis of the proposed method.

Lemma 2.2 [30] *Let $q(u) \in \mathcal{R}^n$ be a monotone mapping of u with respect to $\mathcal{R}^n_+$ and $R \in \mathcal{R}^{n\times n}$ be a positive definite diagonal matrix. For a given $u^k > 0$, if $U_k := diag(u^k_1, u^k_2, \cdots, u^k_n)$ and u^{-1} be an n-vector whose j-th element is $1/u_j$, then the equation*

$$q(u) + R[(u - u^k) + \mu(u^k - U_k^2 u^{-1})] = 0 \tag{2.10}$$

has a unique positive solution u. Moreover, for any $v \geq 0$, we have

$$(v-u)^\top q(u) \geq \tfrac{1+\mu}{2}\left(\|u-v\|_R^2 - \|u^k - v\|_R^2\right) + \tfrac{1-\mu}{2}\|u^k - u\|_R^2. \tag{2.11}$$

In the next theorem we show that α_k is lower bounded away from zero and it is useful for the convergence analysis.

Theorem 2.1 *For given $w^k \in \mathcal{R}^n_{++} \times \mathcal{R}^m_{++} \times \mathcal{R}^l$, let $\tilde{w}^k$ be generated by (2.4)-(2.6), then we have the following*

$$\varphi_k \geq \frac{2-\sqrt{2}}{2}\|w^k - \tilde{w}^k\|_G^2 \tag{2.12}$$

and

$$\alpha_k \geq \frac{2-\sqrt{2}}{2}. \tag{2.13}$$

Proof: It follows from (2.9) that

$$\begin{aligned}
\varphi_k &= \|w^k - \tilde{w}^k\|_M^2 + (\lambda^k - \tilde{\lambda}^k)^T(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)) \\
&= \|x^k - \tilde{x}^k\|_R^2 + \|Ax^k - A\tilde{x}^k\|_H^2 + \|y^k - \tilde{y}^k\|_S^2 + \|By^k - B\tilde{y}^k\|_H^2 + \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2 \\
&\quad +(\lambda^k - \tilde{\lambda}^k)^T(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)).
\end{aligned} \tag{2.14}$$

By using the CauchySchwarz Inequality, we have

$$(\lambda^k - \tilde{\lambda}^k)^T(A(x^k - \tilde{x}^k)) \geq -\frac{1}{2}\left(\sqrt{2}\|A(x^k - \tilde{x}^k)\|_H^2 + \frac{1}{\sqrt{2}}\|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2\right) \tag{2.15}$$

and

$$(\lambda^k - \tilde{\lambda}^k)^T(B(y^k - \tilde{y}^k)) \geq -\frac{1}{2}\left(\sqrt{2}\|B(y^k - \tilde{y}^k)\|_H^2 + \frac{1}{\sqrt{2}}\|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2\right). \tag{2.16}$$

Substituting (2.15) and (2.16) into (2.14), we get

$$
\begin{aligned}
\varphi_k &\geq \frac{2-\sqrt{2}}{2}\left(\|Ax^k - A\tilde{x}^k\|_H^2 + \|By^k - B\tilde{y}^k\|_H^2 + \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2\right) + \|x^k - \tilde{x}^k\|_R^2 + \|y^k - \tilde{y}^k\|_S^2 \\
&\geq \frac{2-\sqrt{2}}{2}\left(\|Ax^k - A\tilde{x}^k\|_H^2 + \|By^k - B\tilde{y}^k\|_H^2 + \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2\right) \\
&\quad +(2-\sqrt{2})\left(\|x^k - \tilde{x}^k\|_R^2 + \|y^k - \tilde{y}^k\|_S^2\right) \\
&= \frac{2-\sqrt{2}}{2}(\|w^k - \tilde{w}^k\|_G^2 + (1-\mu)\|x^k - \tilde{x}^k\|_R^2 + (1-\mu)\|y^k - \tilde{y}^k\|_S^2) \\
&\geq \frac{2-\sqrt{2}}{2}\|w^k - \tilde{w}^k\|_G^2.
\end{aligned}
$$

Therefore, it follows from (2.8) and (2.12) that

$$
\alpha_k \geq \frac{2-\sqrt{2}}{2}
$$

and this completes the proof. $\square$

3 Basic results

In this section, we prove some basic properties, which will be used to establish the sufficient and necessary conditions for the convergence of the proposed method. The following results are due to applying Lemma 2.2 to the LQP systems in the prediction step of the proposed method.

Lemma 3.1 *For given $w^k = (x^k, y^k, \lambda^k) \in \mathcal{R}^n_{++} \times \mathcal{R}^m_{++} \times \mathcal{R}^l$, let $\tilde{w}^k$ be generated by (2.4)–(2.6). Then for any $w^* = (x^*, y^*, \lambda^*) \in \mathcal{W}^*$, we have*

$$
\begin{aligned}
(w^k - w^*)^\top G(w^k - \tilde{w}^k) &\geq \|w^k - \tilde{w}^k\|_G^2 - \mu\|x^k - \tilde{x}^k\|_R^2 - \mu\|y^k - \tilde{y}^k\|_S^2 \\
&\quad +(\lambda^k - \tilde{\lambda}^k)^\top\left(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)\right)
\end{aligned} \tag{3.1}
$$

and

$$
\begin{aligned}
(w_*^k - \tilde{w}^k)^\top D(w^k, \tilde{w}^k) &\geq (w_*^k - w^k)^\top G(w^k - \tilde{w}^k) + \|w^k - \tilde{w}^k\|_G^2 \\
&\quad -\mu\|x^k - \tilde{x}^k\|_R^2 - \mu\|y^k - \tilde{y}^k\|_S^2,
\end{aligned} \tag{3.2}
$$

where

$$
w_*^k = (x_*^k, y_*^k, \lambda_*^k) := P_{\mathcal{W}}[w^k - \alpha_k G^{-1} d(w^k, \tilde{w}^k)]. \tag{3.3}
$$

Proof: Applying Lemma 2.2 to (2.4) by setting $u^k = x^k$, $u = \tilde{x}^k$, $v = x^*$ in (2.11) and

$$
q(u) = f(\tilde{x}^k) - A^\top[\lambda^k - H(A\tilde{x}^k + By^k - b)],
$$

we get

$$\begin{aligned}&(x^* - \tilde{x}^k)^\top \Big\{ f(\tilde{x}^k) - A^T\left[\lambda^k - H(A\tilde{x}^k + By^k - b)\right]\Big\} \\ \geq\;& \tfrac{1+\mu}{2}\Big(\|\tilde{x}^k - x^*\|_R^2 - \|x^k - x^*\|_R^2\Big) + \tfrac{1-\mu}{2}\|x^k - \tilde{x}^k\|_R^2.\end{aligned} \tag{3.4}$$

Recall

$$(x^* - \tilde{x}^k)^\top R(x^k - \tilde{x}^k) = \frac{1}{2}\left(\|\tilde{x}^k - x^*\|_R^2 - \|x^k - x^*\|_R^2\right) + \frac{1}{2}\|x^k - \tilde{x}^k\|_R^2. \tag{3.5}$$

Adding (3.4) and (3.5), we obtain

$$\begin{aligned}&(x^* - \tilde{x}^k)^\top \Big\{(1+\mu)R(x^k - \tilde{x}^k) - f(\tilde{x}^k) + A^\top\tilde{\lambda}^k + A^T HA(x^k - \tilde{x}^k) \\ &-A^T H\left(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)\right)\Big\} \leq \mu\|x^k - \tilde{x}^k\|_R^2.\end{aligned} \tag{3.6}$$

Similarly, applying Lemma 2.2 to (2.5), substituting $u^k = y^k$, $u = \tilde{y}^k$, $v = y^*$ and replacing R, n with S, m, respectively in (2.11) and

$$q(u) = g(\tilde{y}^k) - B^\top[\lambda^k - H(Ax^k + B\tilde{y}^k - b)],$$

we get

$$\begin{aligned}&(y^* - \tilde{y}^k)^\top\{g(\tilde{y}^k) - B^\top[\lambda^k - H(Ax^k + B\tilde{y}^k - b)]\} \\ \geq\;& \tfrac{1+\mu}{2}\Big(\|\tilde{y}^k - y^*\|_S^2 - \|y^k - y^*\|_S^2\Big) + \tfrac{1-\mu}{2}\|y^k - \tilde{y}^k\|_S^2.\end{aligned} \tag{3.7}$$

Recall

$$(y^* - \tilde{y}^k)^\top S(y^k - \tilde{y}^k) = \frac{1}{2}\left(\|\tilde{y}^k - y^*\|_S^2 - \|y^k - y^*\|_S^2\right) + \frac{1}{2}\|y^k - \tilde{y}^k\|_S^2. \tag{3.8}$$

Adding (3.7) and (3.8), we have

$$\begin{aligned}&(y^* - \tilde{y}^k)^\top \Big\{(1+\mu)S(y^k - \tilde{y}^k) - g(\tilde{y}^k) + B^\top\tilde{\lambda}^k + B^T HB(y^k - \tilde{y}^k) \\ &-B^T H(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k))\Big\} \leq \mu\|y^k - \tilde{y}^k\|_S^2,\end{aligned} \tag{3.9}$$

Since (x^*, y^*, λ^*) is a solution of SVI, $\tilde{x}^k \in \mathcal{R}^n_{++}$ and $\tilde{y}^k \in \mathcal{R}^m_{++}$, we have

$$(\tilde{x}^k - x^*)^\top(f(x^*) - A^\top\lambda^*) \geq 0,$$

$$(\tilde{y}^k - y^*)^\top(g(y^*) - B^\top\lambda^*) \geq 0,$$

and

$$Ax^* + By^* - b = 0.$$

Using the monotonicity of f and g, we obtain

$$\begin{pmatrix} \tilde{x}^k - x^* \\ \tilde{y}^k - y^* \\ \tilde{\lambda}^k - \lambda^* \end{pmatrix}^\top \begin{pmatrix} f(\tilde{x}^k) - A^\top\tilde{\lambda}^k \\ g(\tilde{y}^k) - B^\top\tilde{\lambda}^k \\ A\tilde{x}^k + B\tilde{y}^k - b \end{pmatrix} \geq \begin{pmatrix} \tilde{x}^k - x^* \\ \tilde{y}^k - y^* \\ \tilde{\lambda}^k - \lambda^* \end{pmatrix}^\top \begin{pmatrix} f(x^*) - A^\top\lambda^* \\ g(y^*) - B^\top\lambda^* \\ Ax^* + By^* - b \end{pmatrix} \geq 0. \tag{3.10}$$

Adding (3.6), (3.9) and (3.10), we get

$$
\begin{aligned}
& (w^* - \tilde{w}^k)^\top G(w^k - \tilde{w}^k) \\
= \;& (x^* - \tilde{x}^k)^\top((1+\mu)R(x^k - \tilde{x}^k) + A^\top HA(x^k - \tilde{x}^k)) \\
& +(y^* - \tilde{y}^k)^\top((1+\mu)S(y^k - \tilde{y}^k) + B^\top HB(y^k - \tilde{y}^k)) \\
& +(\lambda^* - \tilde{\lambda}^k)^\top(A\tilde{x}^k + B\tilde{y}^k - b) \\
\leq \;& \mu\|x^k - \tilde{x}^k\|_R^2 + (x^* - \tilde{x}^k)^\top A^\top H\left(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)\right) \\
& +(y^* - \tilde{y}^k)^\top B^\top H\left(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)\right) + \mu\|y^k - \tilde{y}^k\|_S^2 \\
= \;& \mu\|x^k - \tilde{x}^k\|_R^2 - (A\tilde{x}^k + B\tilde{y}^k - b)^\top H\left(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)\right) + \mu\|y^k - \tilde{y}^k\|_S^2 \\
= \;& \mu\|x^k - \tilde{x}^k\|_R^2 - (\lambda^k - \tilde{\lambda}^k)^\top\left(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)\right) + \mu\|y^k - \tilde{y}^k\|_S^2
\end{aligned}
\tag{3.11}
$$

where the last equality follows from (2.6). It follows from (3.11) that

$$
\begin{aligned}
& (w^k - w^*)^\top G(w^k - \tilde{w}^k) \\
\geq \;& \|w^k - \tilde{w}^k\|_G^2 - \mu\|x^k - \tilde{x}^k\|_R^2 - \mu\|y^k - \tilde{y}^k\|_S^2 + (\lambda^k - \tilde{\lambda}^k)^\top\left(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)\right)
\end{aligned}
$$

and the first assertion of this lemma is proved.

Similarly as in (3.6) and (3.9), we have

$$
\begin{aligned}
& (x_*^k - \tilde{x}^k)^\top\Big\{(1+\mu)R(x^k - \tilde{x}^k) - f(\tilde{x}^k) + A^\top\tilde{\lambda}^k + A^THA(x^k - \tilde{x}^k) \\
& -A^TH\left(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)\right)\Big\} \leq \mu\|x^k - \tilde{x}^k\|_R^2
\end{aligned}
\tag{3.12}
$$

and

$$
\begin{aligned}
& (y_*^k - \tilde{y}^k)^T\Big\{(1+\mu)S(y^k - \tilde{y}^k) - g(\tilde{y}^k) + B^\top\tilde{\lambda}^k + B^THB(y^k - \tilde{y}^k) \\
& -B^TH(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k))\Big\} \leq \mu\|y^k - \tilde{y}^k\|_S^2.
\end{aligned}
\tag{3.13}
$$

It follows from (3.12) and (3.13) that

$$
\begin{pmatrix} x_*^k - \tilde{x}^k \\ y_*^k - \tilde{y}^k \\ \lambda_*^k - \tilde{\lambda}^k \end{pmatrix}^\top
\begin{pmatrix} ((1+\mu)R + A^THA)(x^k - \tilde{x}^k) - f(\tilde{x}^k) + A^\top\tilde{\lambda}^k - A^TH\left(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)\right) \\ ((1+\mu)S + B^\top HB)(y^k - \tilde{y}^k) - g(\tilde{y}^k) + B^\top\tilde{\lambda}^k - B^\top H(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)) \\ H^{-1}(\lambda^k - \tilde{\lambda}^k) - (A\tilde{x}^k + B\tilde{y}^k - b) \end{pmatrix}
$$

$$
\leq \mu\|x^k - \tilde{x}^k\|_R^2 + \mu\|y^k - \tilde{y}^k\|_S^2,
$$

which implies

$$
(w_*^k - \tilde{w}^k)^\top(G(w^k - \tilde{w}^k) - D(w^k, \tilde{w}^k)) - \mu\|x^k - \tilde{x}^k\|_R^2 - \mu\|y^k - \tilde{y}^k\|_S^2 \leq 0.
$$

By simple manipulation, we obtain

$$
\begin{aligned}
& (w_*^k - \tilde{w}^k)^\top D(w^k, \tilde{w}^k) \\
\geq\ & (w_*^k - \tilde{w}^k)^\top G(w^k - \tilde{w}^k) - \mu\|x^k - \tilde{x}^k\|_R^2 - \mu\|y^k - \tilde{y}^k\|_S^2 \\
=\ & (w_*^k - w^k)^\top G(w^k - \tilde{w}^k) + \|w^k - \tilde{w}^k\|_G^2 - \mu\|x^k - \tilde{x}^k\|_R^2 - \mu\|y^k - \tilde{y}^k\|_S^2
\end{aligned}
$$

and the second assertion of this lemma is proved. □ The following theorem provides a unified framework for proving the convergence of the new algorithm.

Theorem 3.1 *Let $w^* \in \mathcal{W}^*, w^{k+1}(\alpha_k)$ be defined by (2.7) and*

$$\Theta(\alpha_k) := \|w^k - w^*\|_G^2 - \|w^{k+1}(\alpha_k) - w^*\|_G^2, \tag{3.14}$$

then

$$
\begin{aligned}
\Theta(\alpha_k) \ \geq\ & \sigma(\|w^k - w_*^k - \alpha_k(\beta_1 + \beta_2)(w^k - \tilde{w}^k)\|_G^2 \\
& +2\alpha_k(\beta_1 + \beta_2)\varphi_k - \alpha_k^2(\beta_1 + \beta_2)^2\|w^k - \tilde{w}^k\|_G^2).
\end{aligned} \tag{3.15}
$$

Proof: Since $w^* \in \mathcal{W}^*$ and $w_*^k = P_{\mathcal{W}}[w^k - \alpha_k G^{-1} d(w^k, \tilde{w}^k)]$, it follows from (2.3) that

$$\|w_*^k - w^*\|_G^2 \leq \|w^k - \alpha_k G^{-1} d(w^k, \tilde{w}^k) - w^*\|_G^2 - \|w^k - \alpha_k G^{-1} d(w^k, \tilde{w}^k) - w_*^k\|_G^2. \tag{3.16}$$

From (2.7), we get

$$
\begin{aligned}
& \|w^{k+1}(\alpha_k) - w^*\|_G^2 \\
=\ & \|(1-\sigma)(w^k - w^*) + \sigma(w_*^k - w^*)\|_G^2 \\
=\ & (1-\sigma)^2\|w^k - w^*\|_G^2 + \sigma^2\|w_*^k - w^*\|_G^2 + 2\sigma(1-\sigma)(w^k - w^*)^T G(w_*^k - w^*).
\end{aligned}
$$

Using the following identity

$$2(a+b)^\top Gb = \|a+b\|_G^2 - \|a\|_G^2 + \|b\|_G^2$$

for $a = w^k - w_*^k$, $b = w_*^k - w^*$, and (3.16), we obtain

$$\begin{array}{rl} & \|w^{k+1}(\alpha_k) - w^*\|_G^2 \\ = & (1-\sigma)^2\|w^k - w^*\|_G^2 + \sigma^2\|w_*^k - w^*\|_G^2 + \sigma(1-\sigma)\{\|w^k - w^*\|_G^2 \\ & -\|w^k - w_*^k\|_G^2 + \|w_*^k - w^*\|_G^2\} \\ = & (1-\sigma)\|w^k - w^*\|_G^2 + \sigma\|w_*^k - w^*\|_G^2 - \sigma(1-\sigma)\|w^k - w_*^k\|_G^2 \\ \leq & (1-\sigma)\|w^k - w^*\|_G^2 + \sigma\|w^k - \alpha_k G^{-1}d(w^k, \tilde{w}^k) - w^*\|_G^2 \\ & -\sigma\|w^k - \alpha_k G^{-1}d(w^k, \tilde{w}^k) - w_*^k\|_G^2 - \sigma(1-\sigma)\|w^k - w_*^k\|_G^2 \\ \leq & (1-\sigma)\|w^k - w^*\|_G^2 + \sigma\|w^k - \alpha_k G^{-1}d(w^k, \tilde{w}^k) - w^*\|_G^2 \\ & -\sigma\|w^k - \alpha_k G^{-1}d(w^k, \tilde{w}^k) - w_*^k\|_G^2. \end{array} \tag{3.17}$$

Using the definition of $\Theta(\alpha_k)$ and (3.17), we get

$$\begin{array}{rcl} \Theta(\alpha_k) & \geq & \sigma\|w^k - w_*^k\|_G^2 + 2\sigma\alpha_k(w_*^k - w^k)^T d(w^k, \tilde{w}^k) \\ & & +2\sigma\alpha_k(w^k - w^*)^T d(w^k, \tilde{w}^k). \end{array} \tag{3.18}$$

It follows from (3.10) that

$$\begin{array}{rcl} (\tilde{w}^k - w^*)^\top D(w^k, \tilde{w}^k) & \geq & (\tilde{w}^k - w^*)^\top \begin{pmatrix} A^\top H\left(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)\right) \\ B^\top H\left(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)\right) \\ 0 \end{pmatrix} \\ & = & (A\tilde{x}^k + B\tilde{x}^k - b)^\top H\left(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)\right) \\ & = & (\lambda^k - \tilde{\lambda}^k)^\top \left(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)\right). \end{array}$$

Thus,

$$(w^k - w^*)^\top D(w^k, \tilde{w}^k) \geq (w^k - \tilde{w}^k)^\top D(w^k, \tilde{w}^k) + (\lambda^k - \tilde{\lambda}^k)^\top \left(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)\right). \tag{3.19}$$

Applying (3.1) and (3.19) to the last term on the right side of (3.18), we obtain

$$\begin{aligned}\Theta(\alpha_k) \;\geq\;& \sigma\|w^k - w^k_*\|^2_G + 2\sigma\alpha_k(w^k_* - w^k)^\top d(w^k, \tilde{w}^k) \\ & +2\sigma\alpha_k\{\beta_1(w^k - \tilde{w}^k)^\top D(w^k, \tilde{w}^k) + (\beta_1+\beta_2)(\lambda^k - \tilde{\lambda}^k)^\top \big(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)\big) \\ & +\beta_2\|w^k - \tilde{w}^k\|^2_G - \beta_2\mu\|x^k - \tilde{x}^k\|^2_R - \beta_2\mu\|y^k - \tilde{y}^k\|^2_S\} \\ =\;& \sigma\{\|w^k - w^k_*\|^2_G + 2\alpha_k\beta_1(w^k_* - \tilde{w}^k)^\top D(w^k, \tilde{w}^k) \\ & +2\alpha_k\beta_2(w^k_* - w^k)^\top G(w^k - \tilde{w}^k) \\ & +2\alpha_k(\beta_1+\beta_2)(\lambda^k - \tilde{\lambda}^k)^\top \big(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)\big) \\ & +2\alpha_k\beta_2\|w^k - \tilde{w}^k\|^2_G - 2\alpha_k\beta_2\mu\|x^k - \tilde{x}^k\|^2_R - 2\alpha_k\beta_2\mu\|y^k - \tilde{y}^k\|^2_S\}. \end{aligned} \tag{3.20}$$

Applying (3.2) to the second term in the right side of (3.20) and using the notation of φ_k in (2.9), we get

$$\begin{aligned}\Theta(\alpha_k) \;\geq\;& \sigma\{\|w^k - w^k_*\|^2_G + 2\alpha_k(\beta_1+\beta_2)(w^k_* - w^k)^\top G(w^k - \tilde{w}^k) \\ & +2\alpha_k(\beta_1+\beta_2)\big[\|w^k - \tilde{w}^k\|^2_G - \mu\|x^k - \tilde{x}^k\|^2_R - \mu\|y^k - \tilde{y}^k\|^2_S \\ & +(\lambda^k - \tilde{\lambda}^k)^\top \Big(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)\Big)\big]\} \\ =\;& \sigma\{\|w^k - w^k_* - \alpha_k(\beta_1+\beta_2)(w^k - \tilde{w}^k)\|^2_G \\ & -\alpha_k^2(\beta_1+\beta_2)^2\|w^k - \tilde{w}^k\|^2_G + 2\alpha_k(\beta_1+\beta_2)\varphi_k\} \end{aligned}$$

and the theorem is proved. □

From the computational point of view, a relaxation factor $\gamma \in (0, 2)$ is preferable in the correction. We are now in a position to prove the contractive property of the iterative sequence.

Theorem 3.2 *Let $w^* \in \mathcal{W}^*$ be a solution of SVI and let $w^{k+1}(\gamma\alpha_k)$ be generated by (2.7). Then w^k and $\tilde{w}^k$ are bounded, and*

$$\|w^{k+1}(\gamma\alpha_k) - w^*\|^2_G \leq \|w^k - w^*\|^2_G - c\|w^k - \tilde{w}^k\|^2_G, \tag{3.21}$$

where

$$c := \frac{\sigma\gamma(2-\gamma)(2-\sqrt{2})^2}{4} > 0.$$

Proof: It follows from (3.15), (2.12) and (2.13) that

$$
\begin{aligned}
& \|w^{k+1}(\gamma\alpha_k) - w^*\|_G^2 \\
\leq\ & \|w^k - w^*\|_G^2 - \sigma(2\gamma\alpha_k(\beta_1+\beta_2)\varphi_k - \gamma^2\alpha_k^2(\beta_1+\beta_2)^2\|w^k - \tilde{w}^k\|_G^2) \\
=\ & \|w^k - w^*\|_G^2 - \gamma(2-\gamma)(\beta_1+\beta_2)\alpha_k\sigma\varphi_k \\
\leq\ & \|w^k - w^*\|_G^2 - \tfrac{\sigma\gamma(2-\gamma)(2-\sqrt{2})^2}{4}\|w^k - \tilde{w}^k\|_G^2.
\end{aligned}
$$

Since $\gamma \in (0,2)$, we have

$$\|w^{k+1} - w^*\|_G \leq \|w^k - w^*\|_G \leq \cdots \leq \|w^0 - w^*\|_G,$$

and thus, $\{w^k\}$ is a bounded sequence.

It follows from (3.21) that

$$\sum_{k=0}^{\infty} c\|w^k - \tilde{w}^k\|_G^2 < +\infty.$$

which means that

$$\lim_{k\to\infty} \|w^k - \tilde{w}^k\|_G = 0. \tag{3.22}$$

Since $\{w^k\}$ is a bounded sequence, we conclude that $\{\tilde{w}^k\}$ is also bounded. □

4 Convergence of the proposed method

In this section, we prove the global convergence of the proposed method. The following results can be proved by using the technique of Lemma 5.1 and Theorem 5.1 in [2].

Lemma 4.1 *For given $w^k = (x^k, y^k, \lambda^k) \in \mathcal{R}^n_{++} \times \mathcal{R}^m_{++} \times \mathcal{R}^l$, let $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ be generated by (2.4)-(2.6). Then for any $w = (x, y, \lambda) \in \mathcal{W}$, we have*

$$
\begin{aligned}
& (x - \tilde{x}^k)^T \left(f(\tilde{x}^k) - A^T\tilde{\lambda}^k - A^T HA(x^k - \tilde{x}^k) + A^T H \left(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k) \right) \right) \\
\geq\ & (x^k - \tilde{x}^k)^T R\{(1+\mu)x - (\mu x^k + \tilde{x}^k)\}
\end{aligned} \tag{4.1}
$$

and

$$
\begin{aligned}
& (y - \tilde{y}^k)^T \left(g(\tilde{y}^k) - B^T\tilde{\lambda}^k - B^T HB(y^k - \tilde{y}^k) + B^T H \left(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k) \right) \right) \\
\geq\ & (y^k - \tilde{y}^k)^T S\{(1+\mu)y - (\mu y^k + \tilde{y}^k)\}.
\end{aligned} \tag{4.2}
$$

Proof: Applying Lemma 2.2 to prediction step of LQP-ADM (by setting $u^k = x^k$, $u = \tilde{x}^k$, $q(u) = f(\tilde{x}^k) - A^T\left[\lambda^k - H(A\tilde{x}^k + By^k - b)\right]$ and $v = x$ in (2.11)), it follows that

$$\begin{array}{rl} & (x-\tilde{x}^k)^T\Big\{f(\tilde{x}^k) - A^T\left[\lambda^k - H(A\tilde{x}^k + By^k - b)\right]\Big\} \\ \geq & \frac{1+\mu}{2}\Big(\|\tilde{x}^k - x\|_R^2 - \|x^k - x\|_R^2\Big) + \frac{1-\mu}{2}\|x^k - \tilde{x}^k\|_R^2, \end{array}$$

which implies

$$\begin{array}{rl} & (x-\tilde{x}^k)^T\Big(f(\tilde{x}^k) - A^T\tilde{\lambda}^k - A^T HA(x^k - \tilde{x}^k) + A^T H\Big(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)\Big)\Big) \\ \geq & \frac{1+\mu}{2}\Big(\|\tilde{x}^k - x\|_R^2 - \|x^k - x\|_R^2\Big) + \frac{1-\mu}{2}\|x^k - \tilde{x}^k\|_R^2. \end{array}$$

By a simple manipulation, we have

$$\begin{array}{rl} & \frac{1+\mu}{2}\left(\|\tilde{x}^k - x\|_R^2 - \|x^k - x\|_R^2\right) + \frac{1-\mu}{2}\|x^k - \tilde{x}^k\|_R^2 \\ = & (1+\mu)x^T Rx^k - (1+\mu)x^T R\tilde{x}^k - (1-\mu)(\tilde{x}^k)^T Rx^k - \mu\|x^k\|_R^2 + \|\tilde{x}^k\|_R^2 \\ = & (1+\mu)x^T R(x^k - \tilde{x}^k) - (x^k - \tilde{x}^k)^T R(\mu x^k + \tilde{x}^k) \\ = & (x^k - \tilde{x}^k)^T R\big\{(1+\mu)x - \big(\mu x^k + \tilde{x}^k\big)\big\}, \end{array}$$

and the assertion (4.1) is proved. Similarly we can prove the assertion (4.2). □

Now, we are ready to prove the convergence of the proposed method.

Theorem 4.1 *The sequence $\{w^k\}$ generated by the proposed method converges to some w^∞ which is a solution of SVI.*

Proof: It follows from (3.22) that

$$\lim_{k\to\infty}\|x^k - \tilde{x}^k\|_R = 0, \qquad \lim_{k\to\infty}\|y^k - \tilde{y}^k\|_S = 0 \tag{4.3}$$

and

$$\lim_{k\to\infty}\|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}} = \lim_{k\to\infty}\|A\tilde{x}^k + B\tilde{y}^k - b\|_H = 0. \tag{4.4}$$

Moreover, (4.1) and (4.2) imply that

$$\begin{array}{rcl} (x-\tilde{x}^k)^T(f(\tilde{x}^k) - A^T\tilde{\lambda}^k) & \geq & (x^k - \tilde{x}^k)^T R\big\{(1+\mu)x - (\mu x^k + \tilde{x}^k)\big\} \\ & & +(x-\tilde{x}^k)^T\Big(A^T HA(x^k - \tilde{x}^k) - A^T H\Big(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)\Big)\Big) \end{array}$$

and

$$\begin{array}{rcl} (y-\tilde{y}^k)^T(g(\tilde{y}^k) - B^T\tilde{\lambda}^k) & \geq & (y^k - \tilde{y}^k)^T S\big\{(1+\mu)y - (\mu y^k + \tilde{y}^k)\big\} \\ & & +(y-\tilde{y}^k)^T\Big(B^T HB(y^k - \tilde{y}^k) - B^T H\Big(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)\Big)\Big). \end{array}$$

We deduce from (4.3) that

$$\begin{cases} \lim_{k\to\infty}(x-\tilde{x}^k)^T\{f(\tilde{x}^k)-A^T\tilde{\lambda}^k\}\geq 0, & \forall x\in\mathcal{R}^n_{++}, \\ \lim_{k\to\infty}(y-\tilde{y}^k)^T\{g(\tilde{y}^k)-B^T\tilde{\lambda}^k\}\geq 0, & \forall y\in\mathcal{R}^m_{++}. \end{cases} \tag{4.5}$$

Since $\{w^k\}$ is bounded, it has at least one cluster point. Let w^∞ be a cluster point of $\{w^k\}$ and the subsequence $\{w^{k_j}\}$ converges to w^∞, since $\mathcal{W}$ is closed set, we have $w^\infty\in\mathcal{W}$. It follows from (4.4) and (4.5) that (4.5) that

$$\begin{cases} \lim_{j\to\infty}(x-x^{k_j})^T\{f(x^{k_j})-A^T\lambda^{k_j}\}\geq 0, & \forall x\in\mathcal{R}^n_{++}, \\ \lim_{j\to\infty}(y-y^{k_j})^T\{g(y^{k_j})-B^T\lambda^{k_j}\geq 0, & \forall y\in\mathcal{R}^m_{++}, \\ \lim_{j\to\infty}(Ax^{k_j}+By^{k_j}-b)=0. \end{cases}$$

and consequently

$$\begin{cases} (x-x^\infty)^T\{f(x^\infty)-A^T\lambda^\infty\}\geq 0, & \forall x\in\mathcal{R}^n_{++}, \\ (y-y^\infty)^T\{g(y^\infty)-B^T\lambda^\infty\}\geq 0, & \forall y\in\mathcal{R}^m_{++}, \\ Ax^\infty+By^\infty-b=0, \end{cases}$$

which means that w^∞ is a solution of SVI.

Now we prove that the sequence $\{w^k\}$ converges to w^∞. Since

$$\lim_{k\to\infty}\|w^k-\tilde{w}^k\|_G=0, \quad \text{and} \quad \{\tilde{w}^{k_j}\}\to w^\infty,$$

for any $\epsilon>0$, there exists an $l>0$ such that

$$\|\tilde{w}^{k_l}-w^\infty\|_G<\frac{\epsilon}{2} \quad \text{and} \quad \|w^{k_l}-\tilde{w}^{k_l}\|_G<\frac{\epsilon}{2}. \tag{4.6}$$

Therefore, for any $k\geq k_l$, it follows from (3.21) and (4.7) that

$$\|w^k-w^\infty\|_G\leq\|w^{k_l}-w^\infty\|_G\leq\|w^{k_l}-\tilde{w}^{k_l}\|_G+\|\tilde{w}^{k_l}-w^\infty\|_G<\epsilon.$$

This implies that the sequence $\{w^k\}$ converges to w^∞ which is a solution of SVI.

Now we prove that the sequence $\{w^k\}$ converges to w^∞. Since

$$\lim_{k\to\infty}\|w^k-\tilde{w}^k\|_G=0, \quad \text{and} \quad \{\tilde{w}^{k_j}\}\to w^\infty,$$

for any $\epsilon>0$, there exists an $l>0$ such that

$$\left\|\tilde{w}^{k_l}-w^\infty\right\|_G<\frac{\epsilon}{2} \quad \text{and} \quad \left\|w^{k_l}-\tilde{w}^{k_l}\right\|_G<\frac{\epsilon}{2}. \tag{4.7}$$

Therefore, for any $k\geq k_l$, it follows from (3.21) and (4.7) that

$$\|w^k-w^\infty\|_G\leq\|w^{k_l}-w^\infty\|_G\leq\|w^{k_l}-\tilde{w}^{k_l}\|_G+\|\tilde{w}^{k_l}-w^\infty\|_G<\epsilon.$$

This implies that the sequence $\{w^k\}$ converges to w^∞ which is a solution of SVI. □

5 $O(1/t)$ Convergence Rate

In this section, we show by taking $\beta_1 = 1$ and $\beta_2 = 0$ that the proposed method has the $O(1/t)$ convergence rate. Recall that $\mathcal{W}^*$ can be characterized as (see (2.3.2) in pp. 159 of [11])

$$\mathcal{W}^* = \bigcap_{w \in \mathcal{W}} \{\hat{w} \in \mathcal{W} : (w - \hat{w})^T Q(w) \geq 0\}.$$

This implies that $\hat{w}$ is an approximate solution SVI of with the accuracy $\epsilon > 0$ if it satisfies

$$\hat{w} \in \mathcal{W} \quad \text{and} \quad \sup_{w \in \mathcal{W}} \{(w - \hat{w})^T Q(w)\} \leq \epsilon. \tag{5.1}$$

In the rest, our purpose is to show that after t iterations of the proposed method, we can find a $\hat{w} \in \mathcal{W}$ such that (5.1) is satisfied with $\epsilon = O(1/t)$. Since $\beta_1 = 1$ and $\beta_2 = 0$, we have

$$d(w^k, \tilde{w}^k) = \begin{pmatrix} f(\tilde{x}^k) - A^T\tilde{\lambda}^k + A^T H(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)) \\ g(\tilde{y}^k) - B^T\tilde{\lambda}^k + B^T H(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)) \\ A\tilde{x}^k + B\tilde{y}^k - b \end{pmatrix}. \tag{5.2}$$

We introduce some matrices

$$N = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} (1+\mu)R + A^T HA & 0 & 0 \\ 0 & (1+\mu)S + B^T HB & 0 \\ -A & -B & H^{-1} \end{pmatrix}. \tag{5.3}$$

By simple manipulations, we can find that $C = GN$.

Our analysis needs a new sequence defined by

$$\hat{w}^k = \begin{pmatrix} \hat{x}^k \\ \hat{y}^k \\ \hat{\lambda}^k \end{pmatrix} = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \lambda^k - H(Ax^k + By^k - b) \end{pmatrix}. \tag{5.4}$$

Based on (5.3) and (5.4), we easily have a relationship

$$(w^k - \tilde{w}^k) = N(w^k - \hat{w}^k). \tag{5.5}$$

Using (1.3), (5.2) and (5.4), we obtain

$$d(w^k, \tilde{w}^k) = Q(\hat{w}^k). \tag{5.6}$$

Lemma 5.1 *Let $\hat{w}^k$ be defined by (5.4), $w \in \mathcal{W}$ and the matrix C be given in (5.3). Then, we have*

$$(w - \hat{w}^k)^T (Q(\hat{w}^k) - C(w - \hat{w}^k)) \geq -\mu\|x^k - \hat{x}^k\|_R^2 - \mu\|y^k - \hat{y}^k\|_S^2. \tag{5.7}$$

Proof: It follows from (3.6) and (3.9) that

$$
\begin{aligned}
&(x-\tilde{x}^k)^\top\Big\{(1+\mu)R(x^k-\tilde{x}^k)-f(\tilde{x}^k)+A^\top\tilde{\lambda}^k+A^THA(x^k-\tilde{x}^k)\\
&-A^TH\left(A(x^k-\tilde{x}^k)+B(y^k-\tilde{y}^k)\right)\Big\}\le\mu\|x^k-\tilde{x}^k\|_R^2,\quad x\in\mathcal{R}^n_{++}
\end{aligned}
\tag{5.8}
$$

and

$$
\begin{aligned}
&(y-\tilde{y}^k)^\top\Big\{(1+\mu)S(y^k-\tilde{y}^k)-g(\tilde{y}^k)+B^\top\tilde{\lambda}^k+B^THB(y^k-\tilde{y}^k)\\
&-B^TH(A(x^k-\tilde{x}^k)+B(y^k-\tilde{y}^k))\Big\}\le\mu\|y^k-\tilde{y}^k\|_S^2,\quad y\in\mathcal{R}^n_{++}.
\end{aligned}
\tag{5.9}
$$

Then, by using the notation of $\hat{w}^k$ in (5.4), (5.8) and (5.9) can be written as

$$
(x-\hat{x}^k)^\top\Big\{(1+\mu)R(x^k-\hat{x}^k)-f(\hat{x}^k)+A^T\hat{\lambda}^k+A^THA(x^k-\hat{x}^k)\Big\}\le\mu\|x^k-\tilde{x}^k\|_R^2,\quad x\in\mathcal{R}^n_{++}
\tag{5.10}
$$

and

$$
(y-\hat{y}^k)^\top\Big\{(1+\mu)S(y^k-\hat{y}^k)-g(\tilde{y}^k)+B^T\hat{\lambda}^k+B^THB(y^k-\hat{y}^k)\Big\}\le\mu\|y^k-\hat{y}^k\|_S^2,\quad y\in\mathcal{R}^n_{++}.
\tag{5.11}
$$

In addition, it follows from (2.6) and (5.4) that

$$
A\hat{x}^k+B\hat{y}^k-b+H^{-1}(\hat{\lambda}^k-\lambda^k)-A(\hat{x}^k-x^k)-B(\hat{y}^k-y^k)=0.
\tag{5.12}
$$

Combining (5.10)–(5.12), we get

$$
\begin{pmatrix} x^k-\hat{x}^k\\ y^k-\hat{y}^k\\ \lambda^k-\hat{\lambda}^k\end{pmatrix}^\top
\begin{pmatrix} f(\hat{x}^k)-A^\top\hat{\lambda}^k-\left((1+\mu)R+A^THA\right)(x^k-\hat{x}^k)\\ g(\hat{y}^k)-B^\top\hat{\lambda}^k-\left((1+\mu)S+B^\top HB\right)(y^k-\hat{y}^k)\\ A\hat{x}^k+B\hat{y}^k-b-A(x^k-\hat{x}^k)-B(y^k-\hat{y}^k)+H^{-1}(\lambda^k-\hat{\lambda}^k)\end{pmatrix}\ge-\mu\|x^k-\hat{x}^k\|_R^2-\mu\|y^k-\hat{y}^k\|_S^2.
$$

Recall the definition of C in (5.3), we obtain the assertion (5.7). The proof is completed. □

Lemma 5.2 *For given $w^k\in\mathcal{R}^n_{++}\times\mathcal{R}^m_{++}\times\mathcal{R}^l$ and let w^k_* be defined by (3.3). Then, we have*

$$
\gamma\alpha_k(w-\hat{w}^k)^TQ(w)+\frac{1}{2}(\|w-w^k\|_G^2-\|w-w^k_*\|_G^2\ge\frac{1}{2}\gamma(2-\gamma)\alpha_k^2\|w-\tilde{w}^k\|_G^2.
\tag{5.13}
$$

Proof: Since $w_*^k \in \mathcal{R}_{++}^n \times \mathcal{R}_{++}^m \times \mathcal{R}^l$, substituting $w = w_*^k$ in (5.7), we get

$$\begin{aligned}
\gamma\alpha_k(w_*^k - \hat{w}^k)^T Q(\hat{w}^k) &\geq \gamma\alpha_k(w_*^k - \hat{w}^k)^T C(w^k - \hat{w}^k) - \mu\|x^k - \hat{x}^k\|_R^2 - \mu\|y^k - \hat{y}^k\|_S^2 \\
&= \gamma\alpha_k(w^k - \hat{w}^k)^T C(w^k - \hat{w}^k) + \gamma\alpha_k(w_*^k - w^k)^T C(w^k - \hat{w}^k) \\
&\quad -\mu\|x^k - \hat{x}^k\|_R^2 - \mu\|y^k - \hat{y}^k\|_S^2. \\
&= \gamma\alpha_k(w^k - \tilde{w}^k)^T (N^{-1})^T C N^{-1}(w^k - \tilde{w}^k) \\
&\quad +\gamma\alpha_k(w_*^k - w^k)^T C N^{-1}(w^k - \tilde{w}^k) - \mu\|x^k - \tilde{x}^k\|_R^2 - \mu\|y^k - \tilde{y}^k\|_S^2 \\
&= \gamma\alpha_k(w^k - \tilde{w}^k)^T (N^{-1})^T M(w^k - \tilde{w}^k) + \gamma\alpha_k(w_*^k - w^k)^T G(w^k - \tilde{w}^k) \\
&= \gamma\alpha_k\varphi(w^k, \tilde{w}^k) + \gamma\alpha_k(w_*^k - w^k)^T G(w^k - \tilde{w}^k) \\
&\geq \gamma\alpha_k\varphi(w^k, \tilde{w}^k) - \frac{1}{2}\|w^k - w_*^k\|_G^2 - \frac{1}{2}\gamma^2\alpha_k^2\|w^k - \tilde{w}^k\|_G^2 \\
&= \frac{1}{2}\gamma(2-\gamma)\alpha_k^2\|w - \tilde{w}^k\|_G^2 - \frac{1}{2}\|w^k - w_*^k\|_G^2. \qquad (5.14)
\end{aligned}$$

On the other hand, using (3.3) and (5.6), w_*^k is the projection of $w^k - \gamma\alpha_k G^{-1}Q(\hat{w}^k)$ on $\mathcal{W}$, it follows from (2.1) that

$$(w^k - \gamma\alpha_k G^{-1}Q(\hat{w}^k) - w_*^k)^T G(w - w_*^k) \leq 0, \quad \forall w \in \mathcal{W}$$

and consequently

$$\gamma\alpha_k(w - w_*^k)^T Q(\hat{w}^k) \geq (w^k - w_*^k)^T G(w - w_*^k).$$

Using the identity $a^T b = \frac{1}{2}\left(\|a\|^2 - \|a-b\|^2 + \|b\|^2\right)$ to the right hand side of the last inequality, we obtain

$$\gamma\alpha_k(w - w_*^k)^T Q(\hat{w}^k) \geq \frac{1}{2}\left(\|w - w_*^k\|_G^2 - \|w - w^k\|_G^2\right) + \frac{1}{2}\|w^k - w_*^k\|_G^2. \qquad (5.15)$$

Adding (5.14) and (5.15), we get

$$\gamma\alpha_k(w - \hat{w}^k)^T Q(\hat{w}^k) + \frac{1}{2}(\|w - w^k\|_G^2 - \|w - w_*^k\|_G^2 \geq \frac{1}{2}\gamma(2-\gamma)\alpha_k^2\|w - \tilde{w}^k\|_G^2$$

and by using the monotonicity of Q, we obtain(5.13) and the proof is completed. □

Lemma 5.3 *For given $w^k \in \mathcal{R}_{++}^n \times \mathcal{R}_{++}^m \times \mathcal{R}^l$ and let $w^{k+1}(\gamma\alpha_k)$ be generated by (2.7). Then, we have*

$$\gamma\sigma\alpha_k(w - \hat{w}^k)^T Q(w) + \frac{1}{2}(\|w - w^k\|_G^2 - \|w - w^{k+1}(\gamma\alpha_k)\|_G^2 \geq \frac{1}{2}\sigma\gamma(2-\gamma)\alpha_k^2\|w - \tilde{w}^k\|_G^2. \quad (5.16)$$

Proof:

$$\begin{aligned}\|w-w^k\|_G^2-\|w-w^{k+1}(\gamma\alpha_k)\|_G^2 &= \|w^k-w\|_G^2-\|w^k-\sigma(w^k-w_*^k)-w\|_G^2\\ &= 2\sigma\langle w^k-w, w^k-w_*^k\rangle-\sigma^2\|w^k-w_*^k\|_G^2\\ &= 2\sigma\{\|w^k-w_*^k\|_G^2-\langle w-w_*^k, w^k-w_*^k\rangle\}-\sigma^2\|w^k-w_*^k\|_G^2.\end{aligned} \tag{5.17}$$

Using the following identity

$$\langle w-w_*^k, w^k-w_*^k\rangle = \frac{1}{2}\left(\|w_*^k-w\|_G^2-\|w^k-w\|_G^2\right)+\frac{1}{2}\|w^k-w_*^k\|_G^2,$$

implies

$$\|w^k-w_*^k\|_G^2-2\langle w-w_*^k, w^k-w_*^k)\rangle=(\|w^k-w\|_G^2-\|w_*^k-w\|_G^2). \tag{5.18}$$

Substituting (5.18) into (5.17), we obtain

$$\begin{aligned}\|w-w^k\|_G^2-\|w-w^{k+1}(\gamma\alpha_k)\|_G^2 &= \sigma(\|w-w^k\|_G^2-\|w-w_*^k\|_G^2)+\sigma(1-\sigma)\|w^k-w_*^k\|_G^2\\ &\geq \sigma(\|w-w^k\|_G^2-\|w-w_*^k\|_G^2).\end{aligned} \tag{5.19}$$

Substituting (5.19) into (5.13), we obtain (5.16), the required result. □

Now, we are ready to present the $O(1/t)$ convergence rate of the proposed method.

Theorem 5.1 *For any integer $t>0$, we have a $\hat{w}_t\in\mathcal{W}$ which satisfies*

$$(\hat{w}_t-w)^TQ(w)\leq\frac{1}{2\gamma\sigma\Upsilon_t}\|w-w^0\|_G^2,\quad\forall w\in\mathcal{W},$$

where

$$\hat{w}_t=\frac{1}{\Upsilon_t}\sum_{k=0}^{t}\alpha_k\hat{w}^k\quad \textit{and}\quad \Upsilon_t=\sum_{k=0}^{t}\alpha_k.$$

Proof: Summing the inequality (5.16) over $k=0,\cdots,t$, we obtain

$$\left(\left(\sum_{k=0}^{t}\gamma\sigma\alpha_k\right)w-\sum_{k=0}^{t}\gamma\sigma\alpha_k\hat{w}^k\right)^TQ(w)+\frac{1}{2}\|w-w^0\|_G^2\geq 0.$$

Using the notations of Υ_t and $\hat{w}_t$ in the above inequality, we derive

$$(\hat{w}_t-w)^TQ(w)\leq\frac{1}{2\gamma\sigma\Upsilon_t}\|w-w^0\|_G^2,\quad\forall w\in\mathcal{W}.$$

Indeed, $\hat{w}_t\in\mathcal{W}$ because it is a convex combination of $\hat{w}^0,\hat{w}^1,\cdots,\hat{w}^t$. The proof is completed.

□

It follows from (2.13) that

$$\Upsilon_t\geq\frac{2-\sqrt{2}}{2}(t+1).$$

Suppose that for any compact set $\mathcal{D} \subset \mathcal{W}$, let $d = \sup\{\|w - w^0\|_G | w \in \mathcal{D}\}$. For any given $\epsilon > 0$, after most

$$t = [\frac{d^2}{(2-\sqrt{2})\gamma\sigma\epsilon} - 1]$$

iterations, we have

$$(\hat{w}_t - w)^T Q(w) \leq \epsilon, \forall w \in \mathcal{D}.$$

That is, the $O(1/t)$ convergence rate is established in an ergodic sense.

6 Preliminary computational results

In this section we set two examples and applied the proposed method.

6.1 Numerical experiment I

In order to verify the theoretical assertions, we consider the following optimization problem with matrix variables:

$$\min \left\{ \frac{1}{2}\|X - C\|_F^2 : X \in S_+^n \right\}, \tag{6.1}$$

where $\|\cdot\|_F$ is the matrix Fröbenius norm, that is,

$$\|C\|_F = \left(\sum_{i=1}^n \sum_{j=1}^n |C_{ij}|^2 \right)^{1/2},$$

$$S_+^n = \left\{ H \in \mathcal{R}^{n\times n} : H^\top = H, \ H \succeq 0 \right\}.$$

Note that the matrix Fröbenius norm is induced by the inner product

$$\langle A, B \rangle = \text{Trace}(A^\top B).$$

Note that the problem (6.1) is equivalent to the following:

$$\begin{aligned} &\min \tfrac{1}{2}\|X - C\|^2 + \tfrac{1}{2}\|Y - C\|^2 \\ &\text{s.t.} \quad X - Y = 0, \\ &X, Y \in S_+^n, \end{aligned} \tag{6.2}$$

by attaching a Lagrange multiplier $Z \in \mathcal{R}^{n\times n}$ to the linear constraint $X - Y = 0$, the Lagrange function of (6.2) is

$$L(X, Y, Z) = \frac{1}{2}\|X - C\|^2 + \frac{1}{2}\|Y - C\|^2 - \langle Z, X - Y \rangle,$$

which is defined on $S^n_+ \times S^n_+ \times \mathcal{R}^{n\times n}$. If $(X^*, Y^*, Z^*) \in S^n_+ \times S^n_+ \times \mathcal{R}^{n\times n}$ is a KKT point of (6.2), then (6.2) can be converted to the following variational inequality: find $u^* = (X^*, Y^*, Z^*) \in \mathcal{W}$ $= S^n_+ \times S^n_+ \times \mathcal{R}^{n\times n}$ such that

$$\begin{cases} \langle X - X^*, (X^* - C) - Z^*\rangle \geq 0, \\ \langle Y - Y^*, (Y^* - C) + Z^*\rangle \geq 0, \quad \forall\, u = (X, Y, Z) \in \mathcal{W}, \\ X^* - Y^* = 0. \end{cases} \tag{6.3}$$

Problem (6.3) is a special case of (1.2)–(1.3) with matrix variables where $A = I_{n\times n}$, $B = -I_{n\times n}$, $b = 0$, $f(X) = X - C$, $g(Y) = Y - C$ and $\mathcal{W} = S^n_+ \times S^n_+ \times \mathcal{R}^{n\times n}$.

For simplification, we take $R = rI_{n\times n}, S = sI_{n\times n}$ and $H = I_{n\times n}$ where $r > 0$ and $s > 0$ are scalars. In all tests we take $\mu = 0.5$, $\gamma = 1.98$, $\mu = 0.5$, $\sigma = 0.95$, $\beta_1 = 0.5$, $\beta_2 = 0.05$, $C = \text{rand}(n)$ and $(X^0, Y^0, Z^0) = (I_{n\times n}, I_{n\times n}, 0_{n\times n})$ as the initial point in the test, and $r = 1$, $s = 10$ in tables 3–4. The iteration is stopped as soon as

$$\max\left\{\|X^k - \tilde{X}^k\|, \|Y^k - \tilde{Y}^k\|, \|Z^k - \tilde{Z}^k\|\right\} \leq 10^{-6}.$$

All codes were written in Matlab; we compare the proposed method with that in [24]. The iteration numbers, denoted by k, and the computational time for problem (6.1) with different dimensions are given in Tables 1–4.

Tables 1–2 show the efficiency of the proposed method and its superiority to the method of Li [24] in terms of number of iteration and CPU time.

From tables 3–4, we could see that the proposed method works well when β_1 is too large and β_2 is too small. If the parameter β_2 is too large, the iteration numbers and the computational time can increase significantly.

6.2 Traffic equilibrium problems

In this subsection, we apply the proposed method to the traffic equilibrium problems and present corresponding numerical results. We consider a network $[N, L]$ of nodes N and directed links L, which consists of a finite sequence of connecting links with a certain orientation. Let a, b, etc., denote the links, and let p, q, etc., denote the paths. We let ω denote an origin/destination (O/D) pair of nodes of the network and P_ω denotes the set of all paths connecting O/D pair ω. Note that the path-arc incidence matrix and the path-O/D pair incidence matrix, denoted by A and B, respectively, are determined by the given network and

O/D pairs. To see how to convert a traffic equilibrium problem into a variational inequality, we take into account a simple example depicted in Fig.1.

For the given example in Fig.1, the path-arc incidence matrix A and the path-O/D pair incidence matrix B have the following forms:

$$\begin{array}{c} \text{No. link} \quad \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ A = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}, \end{array} \qquad \begin{array}{c} \text{No. O/D pair} \quad \begin{array}{cc} \omega_1 & \omega_2 \end{array} \\ B = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}. \end{array}$$

Let x_p represent the traffic flow on path p and f_a denote the link load on link a, then the arc-flow vector f is given by

$$f = A^\top x. \tag{6.4}$$

Let d_ω denote the traffic amount between O/D pair ω, which must satisfy

$$d_\omega = \sum_{p \in P_\omega} x_p. \tag{6.5}$$

Thus, the O/D pair-traffic amount vector d is given by

$$d = B^\top x. \tag{6.6}$$

Let $t(f) = \{t_a, a \in L\}$ be the vector of link travel costs, which is a function of the link flow. A user travelling on path p incurs a (path) travel cost θ_p. For given link travel cost vector t, the path travel cost vector θ is given by

$$\theta = At(f) \quad \text{and thus} \quad \theta(x) = At(A^\top x). \tag{6.7}$$

Associated with every O/D pair ω, there is a travel disutility $\lambda_\omega(d)$. Since both the path costs and the travel disutilities are functions of the flow pattern x, the traffic network equilibrium problem is to seek the path flow pattern x^* such that

$$x^* \geq 0 \quad (x - x^*)^\top F(x^*) \geq 0, \quad \forall\, x \geq 0 \tag{6.8}$$

where

$$F_p(x) = \theta_p(x) - \lambda_\omega(d(x)), \quad \forall\, \omega, \;\; p \in P_\omega \tag{6.9}$$

and thus

$$F(x) = At(A^\top x) - B\lambda(B^\top x). \tag{6.10}$$

By introducing a positive slack variable $y \geq 0$ and setting $g(y) = 0$ and $B = I$, the problem can be converted into (1.2)-(1.3). The constraints set of problem with link capacity bounds is $S = \{x \in R^n \mid A^\top x \leq b,\ x \geq 0\}$, where b is a given capacity vector. We apply the proposed method to the example taken from [25] (Example 7.4 in [25]), which consisted of 20 nodes, 28 links and 8 O/D pairs. The network is depicted in Figure 2. For this example, there are together 49 paths for the 8 given O/D pairs and hence the dimension of the variable x is 49. Therefore, the path-arc incidence matrix A is a 49×28 matrix and the path-O/D pair incidence matrix B is a 49×8 matrix. The user cost of traversing link a is given in Table 5. The disutility function is given by

$$\lambda_\omega(d) = -m_\omega d_\omega + q_\omega \tag{6.11}$$

and the coefficients m_ω and q_ω in the disutility function of different O/D pairs for this example are given in Table 6.

In all test implementations, we take $x^0 = (1, \dots, 1)^\top, y^0 = (1, \dots, 1)^\top$ and $\lambda^0 = (0, 0, \dots, 0)^\top$ as the starting point, and $\mu = 0.01, \gamma = 1.8, \beta_1 = 0.2, \beta_2 = 1.5, \sigma = 0.95, R = 100I, S = 10I$ and $H = 20I$. For this test problem, the stopping criteria

$$\max\left\{\frac{\|e_x(w^k)\|_\infty}{\|e_x(w^0)\|_\infty}, \|e_y(w^k)\|_\infty, \|e_\lambda(w^k)\|_\infty\right\} \leq \varepsilon, \tag{6.12}$$

where

$$e(w^k) = \begin{pmatrix} e_x(w^k) \\ e_y(w^k) \\ e_\lambda(w^k) \end{pmatrix} = \begin{pmatrix} x^k - P_{\mathcal{R}^n_+}\{x^k - [f(x^k) - A^\top\lambda^k]\} \\ y^k - P_{\mathcal{R}^n_+}\{y^k - [g(y^k) - B^\top\lambda^k]\} \\ Ax^k + By^k - b \end{pmatrix}.$$

We report the numbers of iteration and the CPU time for different capacities and different ε in Tables 7-8. As illustrated in the above, the output vector x is the path-flow, and the link flow vector is $A^T x$. In fact, y^* in the output is referred to as the toll charge on the congested link. For the example with link capacity $b = 40$ we list the optimal link flow and the toll charge in Table 9. Indeed, the link toll charge is greater than zero if and only if the link flow reaches the capacity.

Tables 7–8 show that the proposed method solves the traffic equilibrium problem very efficiently.

7 Conclusions

In this paper, we proposed a new modified logarithmic-quadratic proximal alternating direction method (LQP-ADM) for solving structured variational inequalities. Each iteration of the new LQP-ADM includes a prediction step where a prediction point is obtained by solving series of related systems of nonlinear equations in a parallel wise, and a correction step where the new iterate is generated by searching the optimal step size along a new descent direction. Global convergence of the proposed method is proved under mild assumptions. Some preliminary numerical results are reported to verify the effectiveness of the proposed LQP-ADM in practice. The proposed method converges quite quickly when proper fixed parameters β_1 and β_2 was chosen. However, these proper parameters are unknown beforehand. If the parameter β_2 is large or β_1 is small, the number of iterations could be significantly large. How to choose suitable parameters β_1 and β_2 for different problems is difficult and deserves further research.

Acknowledgement. The authors are very grateful to the referees for their careful reading, comments, and suggestions, which help us improve the presentation of this paper.

References

[1] Auslender A., Teboulle M., Ben-Tiba S.: A logarithmic-quadratic proximal method for variational inequalities, Computational Optimization and Applications 12, 3140 (1999).

[2] Bnouhachem A., Benazza H., Khalfaoui M.: An inexact alternating direction method for solving a class of structured variational inequalities, Applied Mathematics and Computation 219, 78377846 (2013).

[3] Bnouhachem A.: On LQP alternating direction method for solving variational, Journal of Inequalities and Applications 2014(80), 1-15 (2014).

[4] Bnouhachem A., Xu M.H.: An inexact LQP alternating direction method for solving a class of structured variational inequalities, Computer Mathematics and Applications 67, 671680 (2014).

[5] Bnouhachem A. and Ansari Q.H.: A descent LQP alternating direction method for solving variational inequality problems with separable structure, Applied Mathematics and Computation 246, 519-532 (2014).

[6] Bnouhachem A., Hamdi A.: Parallel LQP alternating direction method for solving variational inequality problems with separable structure, Journal of Inequalities and Applications 2014(392), 1-14 (2014).

[7] Chen G., Teboulle M.: A proximal-based decomposition method for convex minimization problems, Mathematical Programming 64, 81101 (1994).

[8] Eckstein J.: Some saddle-function splitting methods for convex programming, Optimization Methods and Software 4, 7583 (1994).

[9] Eckstein J., Bertsekas D.B.: On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, Mathematical Progamming 55, 293318 (1992).

[10] Eckstein J., Fukushima M.: Some reformulation and applications of the alternating directions method of multipliers, in: W.W. Hager et al. (Eds.), Large Scale Optimization: State of the Art, Kluwer Academic Publishers, pp. 115-134 (1994).

[11] Facchinei, F., Pang, J.S.: Finite-Dimensional Variational Inequalities and Complementarity Problems, vol. I and II. Springer Series in Operations Research. Springer, New York (2003).

[12] Fortin M., Glowinski R.: Augmented Lagrangian Methods: Applications to the Solution of Boundary-Valued Problems, North-Holland, Amsterdam (1983).

[13] Gabay D.: Applications of the method of multipliers to variational inequalities, in Augmented Lagrange Methods: Applications to the Solution of Boundary-valued Problems, (eds. M. Fortin and R. Glowinski), Studies in Mathematics and Its Applications, 15, Amsterdam, The Netherlands, pp. 299331 (1983).

[14] Gabay D., B Mercier B.: A dual algorithm for the solution of nonlinear variational problems via finite-element approximations, Compututer and Mathematics with Applications 2, 1740 (1976).

[15] Glowinski R.: Numerical Methods for Nonlinear Variational Problems, Springer-Verlag, New York (1984).

[16] Glowinski R., Le Tallec P.: Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics, SIAM Studies in Applied mathematics, Philadelphia, PA, (1989).

[17] He B.S., Yang H.: Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities, Operation Research Letters 23, 151161 (1998).

[18] He B.S., Liao L.Z., Han D.R., Yang H.: A new inexact alternating directions method for monotone variational inequalities. Mathematical Programming 92, 103118 (2002).

[19] He B.S.: Parallel splitting augmented Lagrangian methods for monotone structured variational inequalities. Computational Optimization and Applications 42, 195212 (2009).

[20] He B.S., Tao M., Yuan X.M.: Alternating direction method with Gaussian back substitution for separable convex programming, SIAM Journal of Optimization 22, 313340 (2012).

[21] Jiang Z.K., Bnouhachem A.: A projection-based prediction-correction method for structured monotone variational inequalities, Applied Mathematics and Computation 202, 747759 (2008).

[22] Jiang Z.K., Yuan X.M.: New parallel descent-like method for sloving a class of variational inequalities, Journal of Optimization Theory and Applications 145, 311-323 (2010).

[23] Kontogiorgis S., Meyer R.R.: A variable-penalty alternating directions method for convex optimization, Mathematical Programming 83, 2953 (1998).

[24] Li M.: A hybrid LQP-based method for structured variational inequalities, International Journal of Computational Mathematics 89(10), 14121425 (2012).

[25] Nagurney A., Zhang D., Projected Dynamical Systems and Variational Inequalities with Applications, Kluwer Academic Publishers, Boston, Dordrecht, London (1996).

[26] Tao M., Yuan X.M.: On the $O(1/t)$ convergence rate of alternating direction method with logarithmic-quadratic proximal regularization, SIAM Journal of Optimization 22(4), 14311448 (2012).

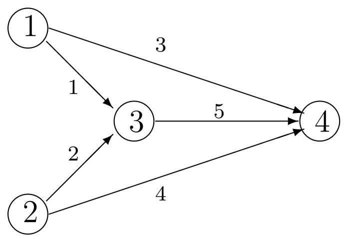
[27] Teboulle M.: Convergence of proximal-like algorithms, SIAM Journal of Optimization 7, 10691083 (1997).

[28] Tseng P.: Alternating projection-proximal methods for convex programming and variational inequalities, SIAM Journal on Optimization 7, 951-965 (1997).

[29] Wang K., Xu L.L., Han D.R.: A new parallel splitting descent method for structured variational inequalities, Journal of Industrial and Management Optimization 10(2), 461-476 (2014).

[30] Yuan X.M., Li M.: An LQP-based decomposition method for solving a class of variational inequalities, SIAM Journal of Optimization 21(4), 13091318 (2011).

List of Figures



O/D Pairs	Path No. & the link on the path
ω_1 : ①→ ④	$p_1 = \{3\}$ $p_2 = \{1, 5\}$
ω_2 : ②→ ④	$p_3 = \{4\}$ $p_4 = \{2, 5\}$

Figure 1: An illustrative example of given directed network and the O/D pairs

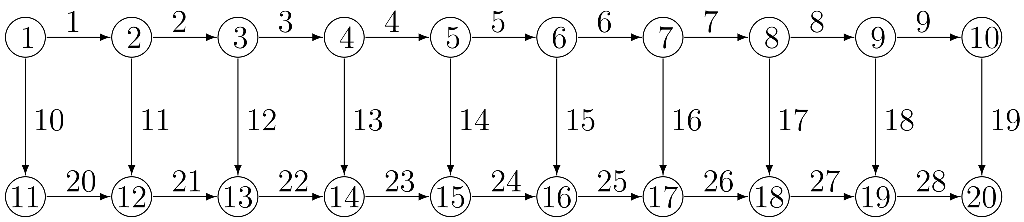


Figure 2: A directed network with 20 nodes and 28 links

List of Tables

Table 1: Numerical results for the problem (6.1) with $r = 0.5$, $s = 5$

Dimension of the problem	The proposed method		The method in [24]	
	k	CPU(Sec.)	k	CPU(Sec.)
100	52	0.98	70	2.32
300	57	5.32	78	7.04
500	60	12.74	82	14.53
700	62	27.12	85	30.98

Table 2: Numerical results for the problem (6.1) with $r = 1$, $s = 10$

Dimension of the problem	The proposed method		The method in [24]	
	k	CPU(Sec.)	k	CPU(Sec.)
100	114	1.13	125	2.54
300	128	7.71	140	8.49
500	134	26.94	147	8.25
700	139	57.53	152	59.27

Table 3: Numerical results of the proposed method for the problem (6.1)with different β_1

Dimension of the problem	$\beta_1 = 0.5$ and $\beta_2 = 0.01$		$\beta_1 = 5$ and $\beta_2 = 0.01$		$\beta_1 = 10$ and $\beta_2 = 0.01$	
	k	CPU(Sec.)	k	CPU(Sec.)	k	CPU(Sec.)
100	114	1.13	110	1.51	109	1.42
300	128	7.71	123	6.75	122	6.71
500	134	26.94	129	27.16	128	26.25
700	139	57.53	133	65.39	132	59.27

Table 4: Numerical results of the proposed method for the problem (6.1) with different β_2

Dimension of the problem	$\beta_1 = 0.5$ and $\beta_2 = 0.01$		$\beta_1 = 0.5$ and $\beta_2 = 0.05$		$\beta_1 = 0.5$ and $\beta_2 = 0.1$	
	k	CPU(Sec.)	k	CPU(Sec.)	k	CPU(Sec.)
100	114	1.13	138	1.13	175	1.94
300	128	7.71	156	8.18	198	10.23
500	134	26.94	163	35.23	207	44.24
700	139	57.53	168	80.55	213	100.22

Table 5: The link traversing cost functions $t_a(f)$ in the example

$t_1(f) = 5 \cdot 10^{-5} f_1^4 + 5f_1 + 2f_2 + 500$	$t_{15}(f) = 3 \cdot 10^{-5} f_{15}^4 + 9f_{15} + 2f_{14} + 200$
$t_2(f) = 3 \cdot 10^{-5} f_2^4 + 4f_2 + 4f_1 + 200$	$t_{16}(f) = 8f_{16} + 5f_{12} + 300$
$t_3(f) = 5 \cdot 10^{-5} f_3^4 + 3f_3 + f_4 + 350$	$t_{17}(f) = 3 \cdot 10^{-5} f_{17}^4 + 7f_{17} + 2f_{15} + 450$
$t_4(f) = 3 \cdot 10^{-5} f_4^4 + 6f_4 + 3f_5 + 400$	$t_{18}(f) = 5f_{18} + f_{16} + 300$
$t_5(f) = 6 \cdot 10^{-5} f_5^4 + 6f_5 + 4f_6 + 600$	$t_{19}(f) = 8f_{19} + 3f_{17} + 600$
$t_6(f) = 7f_6 + 3f_7 + 500$	$t_{20}(f) = 3 \cdot 10^{-5} f_{20}^4 + 6f_{20} + f_{21} + 300$
$t_7(f) = 8 \cdot 10^{-5} f_7^4 + 8f_7 + 2f_8 + 400$	$t_{21}(f) = 4 \cdot 10^{-5} f_{21}^4 + 4f_{21} + f_{22} + 400$
$t_8(f) = 4 \cdot 10^{-5} f_8^4 + 5f_8 + 2f_9 + 650$	$t_{22}(f) = 2 \cdot 10^{-5} f_{22}^4 + 6f_{22} + f_{23} + 500$
$t_9(f) = 10^{-5} f_9^4 + 6f_9 + 2f_{10} + 700$	$t_{23}(f) = 3 \cdot 10^{-5} f_{23}^4 + 9f_{23} + 2f_{24} + 350$
$t_{10}(f) = 4f_{10} + f_{12} + 800$	$t_{24}(f) = 2 \cdot 10^{-5} f_{24}^4 + 8f_{24} + f_{25} + 400$
$t_{11}(f) = 7 \cdot 10^{-5} f_{11}^4 + 7f_{11} + 4f_{12} + 650$	$t_{25}(f) = 3 \cdot 10^{-5} f_{25}^4 + 9f_{25} + 3f_{26} + 450$
$t_{12}(f) = 8f_{12} + 2f_{13} + 700$	$t_{26}(f) = 6 \cdot 10^{-5} f_{26}^4 + 7f_{26} + 8f_{27} + 300$
$t_{13}(f) = 10^{-5} f_{13}^4 + 7f_{13} + 3f_{18} + 600$	$t_{27}(f) = 3 \cdot 10^{-5} f_{27}^4 + 8f_{27} + 3f_{28} + 500$
$t_{14}(f) = 8f_{14} + 3f_{15} + 500$	$t_{28}(f) = 3 \cdot 10^{-5} f_{28}^4 + 7f_{28} + 650$

Table 6: The O/D pairs and the parameters in (6.11) of the example

(O,D) Pair ω	$(1,20)$	$(1,19)$	$(2,17)$	$(4,20)$	$(6,19)$	$(2,20)$	$(2,13)$	$(3,14)$
m_ω	5	6	1	6	10	10	5	4
q_ω	1000	2000	5000	1000	5000	2000	1000	2000
No. of the paths	10	9	6	7	4	9	2	2

Table 7: Numerical results for for different ε with $b = 30$

Different	The proposed method		The method in [30]	
ε	No. It.	CPU(Sec.)	No. It.	CPU(Sec.)
10^{-5}	107	0.68	169	1.12
10^{-6}	125	0.79	197	1.18
10^{-7}	143	0.87	225	1.39
10^{-8}	161	0.96	253	1.15

Table 8: Numerical results for different ε with $b = 40$

Different	The proposed method		The method in [30]	
ε	No. It.	CPU(Sec.)	No. It.	CPU(Sec.)
10^{-5}	98	0.72	154	1.53
10^{-6}	114	0.85	180	1.71
10^{-7}	131	0.99	206	1.85
10^{-8}	148	1.11	231	2.11

Table 9: The optimal link flow and the toll charge on the link when $b = 40$

Link	Flow	Charge	Link	Flow	Charge	Link	Flow	Charge	Link	Flow	Charge
1	0	0	8	32.90	0	15	27.06	0	22	33.95	0
2	12.94	0	9	0	0	16	5.27	0	23	0	0
3	40.00	251.6	10	0	0	17	1.83	0	24	12.94	0
4	12.94	0	11	0	0	18	32.90	0	25	40.00	1245.5
5	0	0	12	33.95	0	19	0	0	26	32.33	0
6	40.00	1254.1	13	27.06	0	20	0	0	27	34.16	0
7	34.73	0	14	12.94	0	21	0	0	28	0	0