Appl. Math. J. Chinese Univ.
2019, 34(4): 446-459

An Augmented Lagrangian based Semismooth Newton Method for a Class of Bilinear Programming Problems

HE Su-xiang   LIU Yan   WANG Chuan-mei

Abstract. This paper proposes a semismooth Newton method for a class of bilinear programming problems (BLPs) based on the augmented Lagrangian, in which the BLPs are reformulated as a system of nonlinear equations with original variables and Lagrange multipliers. Without strict complementarity, the convergence of the method is studied by means of theories of semismooth analysis under the linear independence constraint qualification and strong second order sufficient condition. At last, numerical results are reported to show the performance of the proposed method.

§1 Introduction

Bilinear programming has nonlinearities such that the optimization problem reduces to a linear programming if one of the two variable sets that cause nonlinearities is fixed. And bilinear programming is a subset of nonconvex quadratic programming. Although bilinear programming can be regarded as an extension of linear programming and quadratic programming from its form, bilinear programming has its own wide range of applications in economics, control theory, engineering, management and so on (see [3,7,11,15,16]).

Up until now, the various types of bilinear programming have been studied, such as disjoint bilinear programming, jointly constrained bilinear programming, generalized bilinear programming, bilinear integer programming and so on. And many corresponding methods have been proposed for solving them. For example, the cutting plane technique is a kind of classical approaches for solving bilinear programming, in which the feasible set or objective function is iteratively refined by means of linear inequalities (see [12, 24, 29]). The branch and bound method is also popular for solving bilinear programming, which covers many delicate techniques

Received: 2018-06-12. Revised: 2019-03-12.
MR Subject Classification: 90C26, 90C30.
Keywords: semismooth Newton method, constrained bilinear programming problems, augmented Lagrangian, strict complementarity.
Digital Object Identifier(DOI):https://doi.org/10.1007/s11766-019-3621-5.
Supported by the National Natural Science Foundation of China (No.11671183), the Fundamental Research Funds for the Central Universities (No.2018IB016, 2019IA004, No.2019IB010).

for bounding operation (see [1-3]). Especially, the reformulation-linearization technique is developed in [23] for general bilinear programs, which is a provably convergent branch-and-bound algorithm. And the global optimization algorithm which is a Lagrangian relaxation based approach is suggested in [27] for bilinear programming problems with application in pooling problems.

Motivated by the significance of the augmented Lagrangian method, this paper aims at studying a class of constrained bilinear programming problems of the form below

$$\begin{array}{ll} \min & F(x,y)=c^Tx+d^Ty \\ \text{s.t.} & g_i(x,y)=\alpha_i+\beta_i^Tx+\gamma_i^Ty+x^TH_iy\le 0,\ i=1,\cdots,p, \end{array} \tag{1}$$

where $x\in\Re^n$, $y\in\Re^m$, $c\in\Re^n, d\in\Re^m$, $\alpha_i\in\Re$, $\beta_i\in\Re^n$, $\gamma_i\in\Re^m$, $H_i\in\Re^{n\times m}$, $i=1,\cdots,p$, $n\le m$ and the bilinearities occur in the constrained functions. And problem (1) plays an important role in operations research. The characterizations of linear independent constraint qualification, and the the optimality conditions of problem (1) have the special structure. Moreover, noticing that the augmented Lagrange method is a kind of efficient method for solving constrained optimization problems and the corresponding values of Lagrange multipliers for constraints are important in sensitivity analysis (see [9,13,18,21,22]), we are interested in presenting a novel method for problem (1) by employing the well-known augmented Lagrangian that is a core in the augmented Lagrange method.

In this paper, problem (1) is reformulated as a semismooth system of nonlinear equations with original variables and the Lagrange multipliers based on the favorable properties of the augmented Lagrangian and the special structure of problem (1). In view of the better performance of semismooth Newton method (see [17,19,20,28]), an augmented Lagrangian based semismooth Newton method is presented to solve the system of equations. Furthermore, without strict complementarity, the convergence analysis is developed for the proposed method under the linear independent constraint qualification and the strong second order sufficient condition. Specifically, by applying the theories of semismooth analysis, the sequence of solutions generated by the method is proven to superlinearly converge to the K-K-T solution to problem (1) whenever the controlling parameter in the augmented Lagrangian is larger than a threshold. We can find that the idea of the proposed method is different from that of the augmented Lagrange method, in which the augmented Lagrangian function is minimized and the Lagrange multipliers are updated alternatively, and the linear convergence results are obtained with the strict complementarity condition and without the strict complementarity condition, respectively (see [4,10,26]).

The remainder of this paper is organized as follows. In Section 2, we characterize the K-K-T condition, the linear independent constraint qualification, and the strong second order sufficient condition for problem (1) and recall the basic results from semismooth analysis. Section 3 presents a semismooth Newton method for problem (1) based on the augmented Lagrangian, and carries out the convergence analysis of the developed method without strict complementarity. In Section 4, the preliminary numerical results for the novel method are reported. The concluding remarks are provided in the last section.

§2 Preliminaries

This section serves as a preparation for the convergence analysis of the semismooth Newton method proposed in the next section. This section firstly presents the representations of the K-K-T condition, linear independent constraint qualification and strong second order sufficient condition for problem (1), which are characterized in the specific form corresponding to problem (1). And then some well-known results for semismooth vector-valued functions are recalled.

Let $z = (x^T,\ y^T)^T,\ \xi = (c^T,\ d^T),\ \zeta_i = (\beta_i^T,\ \gamma_i^T)$. Then problem (1) can be written as

$$\begin{aligned} &\min F(z) = \xi z \\ &\text{s.t.}\ \ g_i(z) = \alpha_i + \zeta_i z + x^T H_i y \le 0,\ \ i = 1, \dots, p. \end{aligned}$$

Furthermore, let

$$\tilde{z} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \text{ and } G_i = \begin{pmatrix} 0 & H_i & \beta_i \\ H_i^T & 0 & \gamma_i \\ \beta_i^T & \gamma_i^T & 2\alpha_i \end{pmatrix}.$$

Then problem (1) is simplified as

$$\begin{aligned} \min\ \ &F(z) = \xi z \\ \text{s.t.}\ \ &g_i(z) = \tfrac{1}{2}\tilde{z}^T G_i \tilde{z} \le 0, i = 1, \cdots, p. \end{aligned} \tag{2}$$

Hence the classical Lagrangian function of problem (1) can be expressed as

$$L(x, y, \lambda) = \xi z + \frac{1}{2}\sum_{i=1}^{p} \lambda_i \tilde{z}^T G_i \tilde{z},$$

where λ_i is Lagrange multiplier related to constraint $\frac{1}{2}\tilde{z}^T G_i \tilde{z}$ $(i = 1, \cdots, p)$. And from

$$\nabla g_i(z) = \nabla_z(\frac{1}{2}\tilde{z}^T G_i \tilde{z}) = \begin{pmatrix} \beta_i + H_i y \\ \gamma_i + H_i^T x \end{pmatrix},$$

and

$$\nabla^2 g_i(z) = \nabla_z^2(\frac{1}{2}\tilde{z}^T G_i \tilde{z}) = \begin{pmatrix} 0_{n\times n} & H_i \\ H_i^T & 0_{m\times m} \end{pmatrix}, i = 1, \cdots, p,$$

we get

$$\nabla_z L(x, y, \lambda) = \begin{pmatrix} c \\ d \end{pmatrix} + \textstyle\sum_{i=1}^{p} \lambda_i \begin{pmatrix} \beta_i + H_i y \\ \gamma_i + H_i^T x \end{pmatrix},$$

and

$$\nabla_{zz}^2 L(z, \lambda) = \begin{pmatrix} 0_{n\times n} & \sum_{i=1}^{p} \lambda_i H_i \\ \sum_{i=1}^{p} \lambda_i H_i^T & 0_{m\times m} \end{pmatrix}.$$

Assume that $(\bar{z}, \bar{\lambda}) \in \Re^{n+m} \times \Re^p$ is a K-K-T solution to problem (1), where $\bar{z}$ is a local minimizer of problem (1). Define

$$I_1(\bar{z}, \bar{\lambda}) = \{i | \bar{\lambda}_i > 0, i = 1, \dots, p\} = \{1, \dots, r\}\ (r \le p),$$

$$I_0(\bar{z}, \bar{\lambda}) = \{i | \hat{z}^T G_i \hat{z} = 0, i = 1, \dots, p\} = \{1, \dots, r_1\}\ (r \le r_1 \le p),$$

where $\hat{z} = (\bar{z}^T, 1)^T$. Therefore, the K-K-T condition of problem (1) can be written as

$$\begin{cases} c + \sum_{i=1}^{r} \bar{\lambda}_i(\beta_i + H_i\bar{y}) = 0, \\ d + \sum_{i=1}^{r} \bar{\lambda}_i(\gamma_i + H_i^T\bar{x}) = 0, \\ \hat{z}^T G_i \hat{z} = 0, \ i \in I_0(\bar{z}, \bar{\lambda}), \\ \bar{\lambda}_i = 0, \ i \in \{1, \cdots, p\} \backslash I_1(\bar{z}, \bar{\lambda}). \end{cases}$$

And the linear independence constraint qualification condition of problem (1) is described as the following set of vectors

$$\left\{ \begin{pmatrix} \beta_1 + H_1\bar{y} \\ \gamma_1 + H_1^T\bar{x} \end{pmatrix}, \ldots, \begin{pmatrix} \beta_{r_1} + H_{r_1}\bar{y} \\ \gamma_{r_1} + H_{r_1}^T\bar{x} \end{pmatrix} \right\},$$

which is linearly independent.

Furthermore, the strong second-order sufficient condition of problem (1) is stated as follows. For any $\hat{d} \in \Re^{n+m}$ $(\hat{d} \neq 0)$ satisfying $\begin{pmatrix} \beta_i + H_i\bar{y} \\ \gamma_i + H_i^T\bar{x} \end{pmatrix}^T \hat{d} = 0, \ i = 1, \ldots, r$, it holds that

$$\hat{d}^T \nabla_{zz}^2 L(\bar{z}, \bar{\lambda})\hat{d} = \hat{d}^T \begin{pmatrix} 0_{n\times n} & \sum_{i=1}^{r} \bar{\lambda}_i H_i \\ \sum_{i=1}^{r} \bar{\lambda}_i H_i^T & 0_{m\times m} \end{pmatrix} \hat{d} > 0.$$

For the convenience of statement, we now denote the aforementioned K-K-T condition, the linear independence constraint qualification and the strong second-order sufficient condition as (A), (B) and (C), respectively. Next we list some results for semismooth vector-valued functions.

Let X and Y be two finite dimensional real vector spaces. Let $\Psi : X \to Y$ be a locally Lipschitz continuous function. Then Ψ is almost everywhere F(Fréchet)-differentiable by Rademacher's theorem (see Theorem 3.1.6 in [8]). We denote $\mathcal{D}_\Psi$ by the set of F-differentiable points of Ψ and $\mathcal{J}\Psi(x)$ by the F-derivative of Ψ at x for $x \in \mathcal{D}_\Psi$. Then, the Bouligand subdifferential of Ψ at $x \in X$, denoted by $\partial_B\Psi(x)$, is

$$\partial_B\Psi(x) := \left\{ \lim_{k\to\infty} \mathcal{J}\Psi(x^k) \big| x^k \in \mathcal{D}_\Psi, x^k \to x \right\}.$$

The Clark generalized Jacobian of Ψ at x is the convex hull of $\partial_B\Psi(x)$ introduced in [5], i.e.,

$$\partial\Psi(x) := \text{conv}\{\partial_B\Psi(x)\},$$

where $\text{conv}\{\partial_B\Psi(x)\}$ means the convex hull of $\partial_B\Psi(x)$. The concept of semismoothness was first introduced in [14] and was extended in [20] to vector-valued functions.

Definition 2.1 Suppose that $\Psi : X \to Y$ is a locally Lipschitz continuous function. Ψ is said to be semismooth at $x \in X$ if Ψ is directionally differentiable at x, and for any $\triangle x \in X$ and $V \in \partial\Psi(x + \triangle x)$ with $\triangle x \to 0$,

$$\Psi(x + \triangle x) - \Psi(x) - V(\triangle x) = o(\|\triangle x\|).$$

Furthermore, Ψ is said to be strongly semismooth at $x \in X$ if Ψ is semismooth at x and, for any $\triangle x \in X$ and $V \in \partial\Psi(x + \triangle x)$ with $\triangle x \to 0$,

$$\Psi(x + \triangle x) - \Psi(x) - V(\triangle x) = O(\|\triangle x\|^2).$$

The following lemma on the Bouligand-subdifferential of composite functions are useful in analyzing the properties of system of equations in Section 3, which are proved in Lemma 2.1 of Ref. [25].

Lemma 2.2 *Let $F : X \to Y$ be a continuously differentiable function on an open neighborhood Ξ of $\bar{x} \in X$ and $\Phi : \Xi_Y \subseteq Y \to X'$ be a locally Lipschitz continuous function on an open set Ξ_Y containing $\bar{y} := F(\bar{x})$, where X' is a finite-dimensional real vector space. Suppose that Φ is directionally differentiable at every point in Ξ_Y and that $\mathcal{J}F(\bar{x}) : X \to Y$ is onto. Then it holds that*

$$\partial_B(\Phi * F)(\bar{x}) = \partial_B \Phi(\bar{y})\mathcal{J}F(\bar{x}),$$

where " $*$ " stands for the composite operation.

The following result is Debreu Theorem in Debreu (1952) (see [6]), which will be used in the analysis on properties of the augmented Lagrangian given in the next section.

Lemma 2.3. *Let $A \in \Re^{n,n}$ be a symmetric matrix and $B \in \Re^{r,n}$, Then $By = 0$ $(y \in \Re^n \backslash \{0\})$ implies that $y^T Ay > 0$ if and only if there exists a real number $\bar{M} > 0$ such that $A + MB^TB$ is positive definite for any $M > \bar{M}$.*

The following lemma can be obtained from the definition of semismooth function.

Lemma 2.4. *Define $\phi(a) = [a]_+ = \max\{a, 0\}$, where $a \in \Re$. Then it holds that $\partial\phi(a) = \{0\}$ for $a < 0$; $\partial\phi(a) = [0, 1]$ for $a = 0$; and $\partial\phi(a) = \{1\}$ for $a > 0$.*

§3 Semismooth Newton method and its convergence

This section provides the system of equations based on the augmented Lagrangian of problem (1), whose solution is just the K-K-T solution to problem (1), and presents a semismooth Newton method for the system of equations. Moreover, without strict complementarity, the convergence analysis of the semismooth Newton method is discussed under the assumptions of the three conditions (A), (B) and (C) given in Section 2.

It follows from Ref. [22] that the augmented Lagrangian for problem (1) can be expressed as

$$L_\rho(x, y, \lambda) = L_\rho(z, \lambda) = c^T x + d^T y + \frac{1}{2\rho}(\|[\lambda + \rho g(z)]_+\|^2 - \|\lambda\|^2),$$

where $\rho > 0$ is a controlling parameter, and $g(z) = (g_1(z), \ldots, g_p(z))^T$. And the basic steps in the corresponding augmented Lagrangian method consist of obtaining the solution z^k to $\min_{z\in\Re^{n+m}} L_\rho(z, \lambda^k)$ and updating the current Lagrange multiplier λ^k by $\lambda^{k+1} = [\lambda^k + \rho g(z^k)]_+$ at the kth iteration. Furthermore, according to the K-K-T condition, we can obtain that $[\bar{\lambda} + \rho g(\bar{z})]_+ = \bar{\lambda}$ and

$$\begin{aligned} \nabla_z L_\rho(\bar{z}, \bar{\lambda}) &= \begin{pmatrix} c \\ d \end{pmatrix} + \sum_{i=1}^{p} [\bar{\lambda}_i + \rho g_i(\bar{z})]_+ \begin{pmatrix} \beta_i + H_i \bar{y} \\ \gamma_i + H_i^T \bar{x} \end{pmatrix} \\ &= \nabla_z L(\bar{z}, \bar{\lambda}) = 0. \end{aligned}$$

Hence, finding the K-K-T solution to problem (1) is equivalent to solving the following system of equations:

$$\begin{aligned} &\nabla_z L_\rho(z, \lambda) = 0, \\ &\lambda - [\lambda + \rho g(z)]_+ = 0. \end{aligned} \tag{3}$$

Let

$$\Phi(Z) = \begin{pmatrix} \nabla_z L_\rho(Z) \\ \lambda - [\lambda + \rho g(z)]_+ \end{pmatrix},$$

where $Z = (z^T, \lambda^T)^T$. Then the system (3) implies $\Phi(\bar{Z}) = 0$. And it follows from Ref. [20] that $\Phi(Z)$ is semismooth at (z, λ) with $\lambda + \rho g(z) = 0$ and $\Phi(Z)$ is smooth at (z, λ) with $\lambda + \rho g(z) \neq 0$, which means that $\Phi(Z)$ is semismooth at K-K-T point $(\bar{z}, \bar{\lambda})$ of problem (1) if the strict complementarity does not hold, i.e., $r < r_1$. That is, to obtain K-K-T point $(\bar{z}, \bar{\lambda})$ to problem (1) from (3), we just need to solve the semismooth equation below

$$\Phi(Z) = 0. \tag{4}$$

We now present the corresponding semismooth Newton algorithm for solving system (4), in which the merit function $\theta(Z) = \|\Phi(Z)\|^2$ is used for the line search.

Algorithm 3.1.

Step 1 Choose $\eta \in (0,1)$, $\bar{\zeta} \in (0, \frac{1}{2})$, and $\epsilon \in (0,1)$ being small enough. Let $Z^0 = (z^0, \lambda^0) \in \Re^{n+m} \times \Re^p$ be an any point and set $k := 0$.

Step 2 If $\|\Phi(Z^k)\| < \epsilon$, then stop and the approximate K-K-T solution Z^k of problem (1) is obtained; Otherwise, select $V_k \in \partial\Phi(Z^k)$.

Step 3 Compute the Newton direction d^k by

$$V_k d^k = -\Phi(Z^k).$$

Step 4 Let l_k be the smallest nonnegative integer l satisfying

$$\theta(Z^k) - \theta(Z^k + \eta^l d^k) \geq 2\bar{\zeta}\eta^l \theta(Z^k),$$

and let $\alpha_k = \eta^{l_k}$.

Step 5 Set $Z^{k+1} = Z^k + \alpha_k d^k$, $k := k+1$, and return to Step 2.

We now state and prove two useful propositions below before the convergence of Algorithm 3.1 is developed.

Proposition 3.1. *Suppose that conditions (A)-(C) hold. Then there exists a constant $\bar{\rho} > 0$ such that any element in $\pi_z \partial(\nabla_z L_\rho)(\bar{z}, \bar{\lambda})$ is positive definite when $\rho > \bar{\rho}$, where $\pi_z \partial(\nabla_z L_\rho)(\bar{z}, \bar{\lambda})$ denotes the projection of $\partial(\nabla_z L_\rho)(\bar{z}, \bar{\lambda})$ onto the space $\Re^{n+m}$.*

Proof. From

$$\nabla_z L_\rho(z, \lambda) = \begin{pmatrix} c \\ d \end{pmatrix} + \sum_{i=1}^{p} [\lambda_i + \rho g_i(z)]_+ \begin{pmatrix} \beta_i + H_i y \\ \gamma_i + H_i^T x \end{pmatrix},$$

for any $\triangle z \in \Re^{n+m}$, we have

$$\begin{aligned} &(\pi_z \partial(\nabla_z L_\rho)(z, \lambda))(\triangle z) \\ &\quad = \sum_{i=1}^{p} [\lambda_i + \rho g_i(z)]_+ \begin{pmatrix} 0 & H_i \\ H_i^T & 0 \end{pmatrix} (\triangle z) + \\ &\qquad \rho \sum_{i=1}^{p} \omega_i \begin{pmatrix} \beta_i + H_i y \\ \gamma_i + H_i^T x \end{pmatrix} \begin{pmatrix} \beta_i + H_i y & \gamma_i + H_i^T x \end{pmatrix} (\triangle z), \end{aligned}$$

where $\pi_z\partial(\nabla_z L_\rho)(z,\lambda)$ denotes the projection of $\partial(\nabla_z L_\rho)(z,\lambda)$ onto the space $\Re^{n+m}$, and $\omega_i \in \partial([.]_+)$, $i=1,\dots,p$. Then from Lemma 2.4, we have

$$\begin{aligned}
&(\pi_z\partial(\nabla_z L_\rho)(\bar z,\bar\lambda))(\triangle z)\\
&= \begin{pmatrix} 0 & \sum_{i=1}^r \bar\lambda_i H_i \\ \sum_{i=1}^r \bar\lambda_i H_i^T & 0 \end{pmatrix}(\triangle z) + \rho\sum_{i=1}^r \nabla g_i(\bar z)\nabla g_i(\bar z)^T(\triangle z) + \\
&\quad +\rho\sum_{i=r+1}^{r_1}\bar\omega_i\nabla g_i(\bar z)\nabla g_i(\bar z)^T(\triangle z)\\
&= \nabla_{zz}^2 L(\bar z,\bar\lambda)(\triangle z) + \rho\sum_{i=1}^r \nabla g_i(\bar z)\nabla g_i(\bar z)^T(\triangle z) + \\
&\quad +\rho\sum_{i=r+1}^{r_1}\bar\omega_i\nabla g_i(\bar z)\nabla g_i(\bar z)^T(\triangle z),
\end{aligned}$$

where $\bar\omega_i \in \partial([0]_+) = [0,1]$, $i=r+1,\dots,r_1$.

From condition (C) and Lemma 2.3, one has that there exist $\bar\rho>0$ and $\mu_0>0$ such that whenever $\rho>\bar\rho$, it holds that

$$\nabla_{zz}^2 L(\bar z,\bar\lambda) + \rho\sum_{i=1}^r \nabla g_i(\bar z)\nabla g_i(\bar z)^T \succeq \mu_0 I.$$

For any $U_\rho \in \pi_z\partial(\nabla_z L_\rho)(\bar z,\bar\lambda)$,

$$U_\rho \succeq \nabla_{zz}^2 L(\bar z,\bar\lambda) + \rho\sum_{i=1}^r \nabla g_i(\bar z)\nabla g_i(\bar z)^T,$$

so we have $U_\rho \succeq \mu_0 I$ for any $\bar\omega_i\in[0,1]$ and $\rho>\bar\rho$. That is, the conclusion holds. □

The next proposition demonstrates the nonsingularity of every element in the Clarke generalized differential of $\Phi(Z)$ at $\bar Z$.

Proposition 3.2. *Suppose that conditions (A)-(C) hold. Then any element $\bar V_\rho \in \partial\Phi(\bar Z)$ is nonsingular whenever $\rho>\bar\rho$, where $\bar\rho$ is defined in Proposition 3.1.*

Proof. For any element $V_\rho(Z)\in\partial\Phi(Z)$, by computing, we obtain

$$V_\rho(Z) = \begin{pmatrix} \nabla_{zz}^2 L(z,\lambda) + \rho\sum_{i=1}^p \theta_i\nabla g_i(z)\nabla g_i(z)^T & \theta_1\nabla g_1(z) & \cdots & \theta_p\nabla g_p(z) \\ -\rho\theta_1\nabla g_1(z)^T & 1-\theta_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\rho\theta_p\nabla g_p(z)^T & 0 & \cdots & 1-\theta_p \end{pmatrix},$$

where $\theta_i\in\partial([.]_+)$, $\nabla g_i(z) = \begin{pmatrix} \beta_i + H_i y \\ \gamma_i + H_i^T x\end{pmatrix}$, $i=1,\dots,p$. Let $\nabla g_{(r)}(z) = (\nabla g_1(z),\dots,\nabla g_r(z))$ and $\nabla g_{(r+1,r_1)}(z) = (\nabla g_{r+1}(z),\dots,\nabla g_{r_1}(z))$. Then it follows from Lemma 2.4 that for any element $\bar V_\rho\in\partial\Phi(\bar Z)$,

$$\bar V_\rho = \begin{pmatrix} A_1 & A_2 & A_3 & 0 \\ -\rho A_2^T & 0 & 0 & 0 \\ -\rho A_3^T & 0 & D & 0 \\ 0 & 0 & 0 & I \end{pmatrix},$$

where $A_1 \in \pi_z\partial(\nabla_z L_\rho)(\bar{z},\bar{\lambda})$, $A_2 = \nabla g_{(r)}(\bar{z})$, $A_3 = \nabla g_{(r+1,r_1)}(\bar{z})\text{diag}_{r+1\leq i\leq r_1}(\bar{\theta}_i)$, $D = \text{diag}_{r+1\leq i\leq r_1}(1-\bar{\theta}_i)$, $I \in \Re^{(p-r_1)\times(p-r_1)}$ is an identity, and $\bar{\theta}_i \in [0,1]$, $i = r+1,\ldots,r_1$.

We now consider the partition of the matrix $\bar{V}_\rho$ in the form below

$$\bar{V}_\rho = \begin{pmatrix} A_\rho & 0 \\ 0 & I \end{pmatrix},$$

where

$$A_\rho = \begin{pmatrix} A_1 & A_2 & A_3 \\ -\rho A_2^T & 0 & 0 \\ -\rho A_3^T & 0 & D \end{pmatrix}.$$

It follows from the property of partitioned matrix that $\bar{V}_\rho$ is nonsingular if A_ρ is nonsingular. Hence, next we devote to proving the nonsingularity of A_ρ from the two cases below.

Case 1. Suppose that $\bar{\theta}_i \neq 1$ for any $i \in \{r+1,\ldots,r_1\}$, which implies that the diagonal matrix D is nonsingular and positive definite. Suppose that $u = (u_1^T, u_2^T, u_3^T)^T \in \Re^{m+n+r_1}$, where $u_1 \in \Re^{m+n}$, $u_2 \in \Re^r$, and $u_3 \in \Re^{r_1-r}$. Let $A_\rho u = 0$, i.e.,

$$A_\rho u = \begin{pmatrix} A_1 & A_2 & A_3 \\ -\rho A_2^T & 0 & 0 \\ -\rho A_3^T & 0 & D \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = 0,$$

which means that

$$A_1u_1 + A_2u_2 + A_3u_3 = 0, \tag{5}$$
$$-\rho A_2^T u_1 = 0, \tag{6}$$
$$-\rho A_3^T u_1 + Du_3 = 0. \tag{7}$$

According to the singularity of D, the formula (7) can be written as

$$u_3 = \rho D^{-1} A_3^T u_1. \tag{8}$$

Multiply both sides of formula (5) on left by u_1^T , we have

$$u_1^T A_1 u_1 + u_1^T A_2 u_2 + u_1^T A_3 u_3 = 0.$$

Therefore, it follows from formula (6) and formula (8) that

$$u_1^T A_1 u_1 + \rho(A_3^T u_1)^T D^{-1}(A_3^T u_1) = 0. \tag{9}$$

Since A_1 is positive definite from Propositions 3.1 whenever $\rho > \bar{\rho}$ and D^{-1} is positive definite, we have

$$u_1^T A_1 u_1 \geq 0,\ (A_3^T u_1)^T D^{-1}(A_3^T u_1) \geq 0.$$

By formula (9), it holds that $u_1 = 0$ and $A_3^T u_1 = 0$. Consequently, from formula (8), we have $u_3 = 0$.

Hence, formula (5) reduces to $A_2u_2 = 0$. Furthermore, the linear independence constraint qualification condition means that A_2 is full of column rank, so we have $u_2 = 0$. That is, the matrix A_ρ is nonsingular.

Case 2. Suppose that there exists $j \in \{r+1,\ldots,r_1\}$ such that $\bar{\theta}_j = 1$. For convenience of statement, we might as well assume that $j = r+1$, i.e., $\bar{\theta}_{r+1} = 1$. Hence, A_ρ can be expressed as

$$A_\rho = \begin{pmatrix} A_1 & A_2 & \nabla g_{r+1}(\bar{z}) & A_4 \\ -\rho A_2^T & 0 & 0 & 0 \\ -\rho \nabla g_{r+1}(\bar{z})^T & 0 & 0 & 0 \\ -\rho A_4^T & 0 & 0 & \hat{D} \end{pmatrix},$$

where $A_4 = (\nabla g_{r+2}(\bar{z}), \ldots, \nabla g_{r_1}(\bar{z}))\text{diag}_{r+2\le i\le r_1}(\bar{\theta}_i)$, $\hat{D} = \text{diag}_{r+2\le i\le r_1}(1-\bar{\theta}_i)$.

Suppose that $u = (u_1^T, u_2^T, u_3, u_4^T)^T \in \Re^{m+n+r_1}$, where $u_1 \in \Re^{m+n}$, $u_2 \in \Re^r$, $u_3 \in \Re^1$ and $u_3 \in \Re^{r_1-(r+1)}$. Let $A_\rho u = 0$. That is,

$$A_1u_1 + A_2u_2 + \nabla g_{r+1}(\bar{z})u_3 + A_4u_4 = 0, \tag{10}$$

$$-\rho A_2^T u_1 = 0, \tag{11}$$

$$-\rho \nabla g_{r+1}(\bar{z})^T u_1 = 0, \tag{12}$$

$$-\rho A_4^T u_1 + \hat{D}u_4 = 0. \tag{13}$$

Since $\hat{D}$ is nonsingular and positive definite from the definition of $\hat{D}$, it follows from the formula (13) that

$$u_4 = \rho \hat{D}^{-1} A_4^T u_1. \tag{14}$$

Multiplying both sides of formula (10) on left by u_1^T , we have

$$u_1^T A_1 u_1 + u_1^T A_2 u_2 + u_1^T \nabla g_{r+1}(\bar{z})u_3 + u_1^T A_4 u_4 = 0.$$

Combined formula (11), formula (12) and formula (14), it holds that

$$u_1^T A_1 u_1 + \rho (A_4^T u_1)^T \hat{D}^{-1} (A_4^T u_1) = 0. \tag{15}$$

Since A_1 is positive definite from Propositions 3.1 whenever $\rho > \bar{\rho}$ and $\hat{D}^{-1}$ is positive definite, we have

$$u_1^T A_1 u_1 \ge 0,\ (A_4^T u_1)^T D^{-1} (A_4^T u_1) \ge 0.$$

By formula (15), it holds that $u_1 = 0$ and $A_4^T u_1 = 0$. Further, from formula (14), we have $u_4 = 0$.

Hence, formula (10) reduces to $A_2u_2 + \nabla g_{r+1}(\bar{z})u_3 = 0$, i.e.,

$$(A_2\ \nabla g_{r+1}(\bar{z})) \begin{pmatrix} u_2 \\ u_3 \end{pmatrix} = 0.$$

Moreover, the linear independence constraint qualification condition means that $(A_2\ \nabla g_{r+1}(\bar{z}))$ is full of column rank, so we have $u_2 = 0$, and $u_3 = 0$. That is, the matrix A_ρ is nonsingular whenever $\rho > \bar{\rho}$.

Suppose that there exist more than one index in $\{r+1, \ldots, r_1\}$, say $j_i \in \{r+1, \ldots, r_1\}$ such that $\bar{\theta}_{j_i} = 1$, where $i \in \{1, \ldots, r_1 - r\}$. Then the conclusion that the matrix A_ρ is nonsingular whenever $\rho > \bar{\rho}$ can also be obtained by the similar analysis to the above proof process.

The proof is completed. □

Corollary 3.1. If conditions (A)-(C) hold, then there exist $\epsilon > 0$ and positive constant c such that for any $Z \in S(\bar{Z}, \epsilon)$ and $V_\rho \in \partial\Phi(Z)$, V_ρ is nonsingular and $\|V_\rho^{-1}\| \le c$ whenever $\rho > \bar{\rho}$, where $S(\bar{Z}, \epsilon) = \{Z \in \Re^{n+m+p} | \|Z - \bar{Z}\| \le \epsilon\}$.

Proof. From Proposition 3.2, any element $\bar{V}_\rho \in \partial\Phi(\bar{Z})$ is nonsingular under conditions (A)-(C) if $\rho > \bar{\rho}$. Hence the conclusion is true by Lemma 2.6 in Ref. [20]. □

Corollary 3.1 shows that Algorithm 3.1 is well defined if Z^0 is close to $\bar{Z}$ enough. Furthermore, the following theorem indicates that the sequence generated by Algorithm 3.1 converges to the K-K-T solution $\bar{Z}$ to problem (1), whose proof can be regarded as an application of Theorem 4.3 in Ref. [19].

Theorem 3.1. Suppose that conditions (A)-(C) hold. If the sequence $\{Z^k\}$ is generated by Algorithm 3.1, then $\{Z^k\}$ converges to the K-K-T solution $\bar{Z}$ superlinearly and α^k eventually becomes 1 whenever $\rho > \bar{\rho}$.

Proof. Firstly, we need to prove that for some $\delta \in (0,1)$, there exists $\hat{\epsilon} > 0$ such that for any $Z \in S(\bar{Z}, \hat{\epsilon})$, it holds that

$$\|Z + d - \bar{Z}\| \leq \delta\|Z - \bar{Z}\|, \text{ and } \|\Phi(x+d)\| \leq \delta\|\Phi(x)\|, \tag{16}$$

where d is the solution to $V_\rho d = -\Phi(Z)$ and $V_\rho \in \partial\Phi(Z)$.

By Corollary 3.1, we know that there exists $\epsilon > 0$ such that $d = -V_\rho^{-1}\Phi(Z)$ is well defined for $Z \in S(\bar{Z}, \epsilon)$ whenever $\rho > \bar{\rho}$. Since $\Phi(Z)$ is semismooth, for $\epsilon > 0$ being small enough, we have

$$\Phi(Z) - \Phi(\bar{Z}) - \Phi'(\bar{Z}; Z - \bar{Z}) = o(\|Z - \bar{Z}\|), \tag{17}$$

$$V_\rho(Z - \bar{Z}) - \Phi'(\bar{Z}; Z - \bar{Z}) = o(\|Z - \bar{Z}\|), \tag{18}$$

where $\Phi'(\bar{Z}; Z - \bar{Z})$ is the directional derivative of $\Phi(Z)$ at $\bar{Z}$ in the direction $Z - \bar{Z}$ (see [5]). Then it follows from (17), (18) and Corollary 3.1 that

$$\begin{aligned} & \|Z + d - \bar{Z}\| \\ = & \|Z - \bar{Z} - V_\rho^{-1}\Phi(Z)\| \\ = & \|V_\rho^{-1}(V_\rho(Z - \bar{Z}) - \Phi(Z))\| \\ \leq & \|V_\rho^{-1}\| \left[\|V_\rho(Z - \bar{Z}) - \Phi'(\bar{Z}; Z - \bar{Z})\| + \|\Phi(Z) - \Phi(\bar{Z}) - \Phi'(\bar{Z}; Z - \bar{Z})\|\right] \\ = & o(\|Z - \bar{Z}\|), \end{aligned}$$

which implies that for some $\delta \in (0,1)$, there exists $\bar{\epsilon} \in (0, \epsilon)$ such that for any $Z \in S(\bar{Z}, \bar{\epsilon})$, one gets

$$\|Z + d - \bar{Z}\| \leq \delta\|Z - \bar{Z}\|. \tag{19}$$

By (17), for the above $\delta \in (0,1)$, there exists $\hat{\epsilon} \in (0, \bar{\epsilon})$ such that if $Z \in S(\bar{Z}, \hat{\epsilon})$, we have

$$\|\Phi(Z) - \Phi(\bar{Z}) - \Phi'(\bar{Z}; Z - \bar{Z}) \leq \delta\|Z - \bar{Z}\|. \tag{20}$$

For any $Z \in S(\bar{Z}, \hat{\epsilon})$, by (19) and Corollary 3.1, we obtain

$$\begin{aligned} \|Z - \bar{Z}\| & \leq \|Z + d - \bar{Z}\| + \|d\| \\ & \leq \delta\|Z - \bar{Z}\| + c\|\Phi(Z)\|, \end{aligned}$$

which means that

$$\|Z - \bar{Z}\| \leq \frac{c}{1-\delta}\|\Phi(Z)\|. \tag{21}$$

Since $\Phi(Z)$ is semismooth at $\bar{Z}$, one has

$$\|\Phi'(\bar{Z}; Z + d - \bar{Z})\| \leq L\|Z + d - \bar{Z}\|,$$

where L is the Lipschitz constant of $\Phi(Z)$ around $\bar{Z}$. And from (19), (20) and (21), it holds that

$$\begin{aligned}\|\Phi(Z+d)\| &\leq \|\Phi'(\bar{Z}, Z+d-\bar{Z})\| + \delta\|Z+d-\bar{Z}\| \\ &\leq (L+\delta)\|Z+d-\bar{Z}\| \\ &\leq (L+\delta)\delta\|Z-\bar{Z}\| \\ &\leq \frac{c\delta(L+\delta)}{1-\delta}\|\Phi(Z)\|,\end{aligned}$$

which means that if $Z \in S(\bar{Z}, \hat{\epsilon})$, it holds that $\|\Phi(Z+d)\| \leq \delta\|\Phi(Z)\|$ for $\delta \in (0,1)$ being small enough whenever $\rho > \bar{\rho}$. That is, (16) is true.

We now turn to discuss the convergence of the sequence $\{Z^k\}$. It follows from (16) that there exists $\tilde{\epsilon} \in (0, \hat{\epsilon})$ such that for some $Z^{\bar{k}}$ with $\|Z^{\bar{k}} - \bar{Z}\| \leq \tilde{\epsilon}$, it is true that

$$\|Z^{\bar{k}} + d^{\bar{k}} - \bar{Z}\| \leq \frac{1}{2}\|Z^{\bar{k}} - \bar{Z}\|, \tag{22}$$

$$\|\Phi(Z^{\bar{k}} + d^{\bar{k}})\| \leq \sqrt{1-2\zeta}\|\Phi(Z^{\bar{k}})\|,$$

where $d^{\bar{k}} \in \partial_B\Phi(Z^{\bar{k}})$.

Hence,

$$\begin{aligned}\theta(Z^{\bar{k}} + d^{\bar{k}}) &= \|\Phi(Z^{\bar{k}} + d^{\bar{k}})\|^2 \\ &\leq (1-2\zeta)\|\Phi(Z^{\bar{k}})\|^2 \\ &= (1-2\zeta)\theta(Z^{\bar{k}}).\end{aligned}$$

That is,

$$\theta(Z^{\bar{k}}) - \theta(Z^{\bar{k}} + d^{\bar{k}}) \geq 2\zeta\theta(Z^{\bar{k}}).$$

Then by Step 4 of Algorithm 3.1, we have $\alpha^{\bar{k}} = 1$. Let $Z^{\bar{k}+1} = Z^{\bar{k}} + d^{\bar{k}}$. In view of (22), one has

$$\|Z^{\bar{k}+1} - \bar{Z}\| \leq \frac{1}{2}\|Z^{\bar{k}} - \bar{Z}\| \leq \|Z^{\bar{k}} - \bar{Z}\| \leq \tilde{\epsilon}.$$

By induction of the above arguments, for any $k \geq \bar{k}$, it holds that

$$\|Z^{k+1} - \bar{Z}\| \leq \frac{1}{2}\|Z^k - \bar{Z}\| \leq \|Z^k - \bar{Z}\| \leq \tilde{\epsilon}, \tag{23}$$

and

$$\alpha^k = 1, \tag{24}$$

where $Z^{k+1} = Z^k + d^k$, and $V_\rho^k d^k = -\Phi(Z^k)$ ($V_\rho^k \in \partial\Phi(Z^k)$.

From condition (B), we know that $\bar{Z}$ is the unique solution of $\Phi(Z) = 0$. And by Step 4 of Algorithm 3.1, $\{\theta(Z^k)\}$ converges to zero. Hence, the sequence $\{Z^k\}$ generated by algorithm 3.1 converges to $\bar{Z}$ superlinearly. Combined (23) and (24), the conclusion is drawn. □

§4 Numerical experiments

We compile program in Matlab language based on Algorithm 3.1 and test some randomly generated problems. The numerical experiments are implemented in the Matlab R2014a running environment on the computer with processor of Intel CORE$^{\text{TM}}$i3-2310M@2.10GHz, and memory capacity of 2 Gb.

In the experiments, we set $\rho = 10^4$, $\eta = 0.1$ and $\bar{\zeta} = 0.3$; the stopping precision is set as $\epsilon = 10^{-5}$ in Step 2 of Algorithm 3.1; the initial point Z^0 is chosen to be the vector whose entries are all ones; and the matrices and vectors in problem (1) are generated randomly by rand function in Matlab. We report the numerical results for seven problems in Table 4.1, in which n, m, p, IT, FN, θ^0 and θ^* represent the dimensional number of x, the dimensional number of y, the number of constraints, the number of iterations, the number of $\theta(Z)$ evaluations, the initial value of $\theta(Z)$ and the final value of $\theta(Z)$, respectively.

Table 4.1. Numerical results

(n, m, p)	IT	FN	θ^0	θ^*
(50,80,40)	14	16	1.2106e+03	7.2123e-06
(80,100,20)	20	21	4.6000e+04	3.4532e-06
(100,50,40)	27	28	3.0058e+05	4.2468e-06
(150,170,30)	35	39	2.9940e+06	2.3089e-06
(180,180,60)	47	52	3.7721e+06	1.6328e-06
(200,240,100)	63	65	2.1137e+07	2.3875e-06
(260,300,200)	81	97	2.7083e+08	3.5213e-06
(320,350,270)	99	105	3.2301e+06	3.1208e-06

Remark. From the preliminary numerical results shown in Table 4.1, we know that the proposed method is feasible and promising for solving problem (1).

§5 Conclusions

A semismooth Newton method for solving a class of bilinear programming problems is explored based on the well-known augmented Lagrangian in this paper. The particular forms of the optimality conditions are characterized according to the special structure of this class problems. By means of the favourable properties of the augmented Lagrangian, this class of bilinear programming problems is transformed to a semismooth system of nonlinear equations, and a semismooth Newton algorithm is presented for solving this system. Without strict complementarity, the semismooth Newton method is proven to be superlinearly convergent under the linear independent constraint qualification and strong second order sufficient condition whenever the controlling parameter in the augmented Lagrangian is larger than a threshold. The preliminary numerical experiments are conducted to demonstrate the effectiveness of the proposed method. In our future work, the program need to be improved to enhance the performance of the method and we also consider to apply this method in some practical problems.

References

[1] S Alarie, C Audet, B Jaurnard, G Savard. *Concavity cuts for disjoint bilinear programming,* Math Program, 2001, 90: 373-398.

[2] F A Al-Khayyal, J E Falk. *Jointly constrained biconvex programming,* Math of Oper Res, 1983, 8: 273-286.

[3] R As′ad, K Demirli. *A bilinear programming model and a modified branch-and-bound algorithm for production planning in steel rolling mills with substitutable demand,* Int J of Prod Res, 2011, 49: 3731-3749.

[4] D P Bertsekas. *Constrained optimization and Lagrange multiplier methods,* 1982, Academic Press, New York.

[5] F H Clarke. *Optimization and nonsmooth analysis,* 1983, Wiley, New York.

[6] G Debreu. *Definite and semidefinite quadratic forms,* Econometrica, 1952, 20: 295-300.

[7] G Erbeyoğlu, Ü Bilge. *PSO-based and SA-based metaheuristics for bilinear programming problems: an application to the pooling problem,* J. Heuristics, 2016, 22: 147-179.

[8] H Federer. *Geometric measure theory,* 1969, Berlin, Heidelberg, New York: Springer-Verlag.

[9] M R Hestenes. *Multiplier and gradient method,* J. Optim Theory Appl, 1969, 4: 303-320.

[10] K Ito, K Kunisch. *The augmented method for equality and inequality constraints in Hilbert spaces,* Math Program, 1990, 46: 341-360.

[11] H Konno. *Bilinear programming, Part II: Applications of bilinear programming,* Technical Report, No. 71-10, 1971, Operations Research House, Stanford University (Stanford, CA).

[12] H Konno. *A cutting plane algorithm for solving bilinear programs,* Math Program, 1976, 11: 14-27.

[13] A Kowalewski, I Lasiecka, J Sokolowski. *Sensitivity analysis of infinite order hyperbolic optimal control problems,* Comput Optim Appl, 2012, 52: 147-179.

[14] R Mifflin. *Semismooth and semiconvex functions in constrained optimization,* SIAM J Control Optim, 1977, 15: 957-972.

[15] A Nahapetyan, P Pardalos. *Adaptive dynamic cost updating procedure for solving fixed charge network flow problems,* Compt Optim Appl, 2008, 39: 37-50.

[16] H Osman, K Demirli. *A bilinear goal programming model and a modified Benders decomposition algorithm for supply chain reconfiguration and supplier selection,* Int J Prod Econ, 2010, 124: 97-105.

[17] J Pang. *Newton's method for B-Differentiable equations,* Math of Oper Res, 1990, 15: 311-341.

[18] M J D Powell. *A method for nonlinear constraints in minimization problems, in: R. Fletcher (Ed.), Optimization,* 1969, Academic Press, New York, pp. 283-298.

[19] L Qi. *Convergence analysis of some algorithms for solving nonsmooth equations,* Math of Oper Res, 1993, 18: 227-245.

[20] L Qi, J Sun. *A nonsmooth version of Newton's method,* Math Program, 1993, 58: 353-367.

[21] R T Rockafellar. *The multiplier method of Hestenes and Powell applied to convex programming,* Math Program, 1973, 12: 555-562.

[22] R T Rockafellar. *Proximal subgradients, marginal values, and augmented Lagrangians in nonconvex optimization,* Math of Oper Res, 1981, 6: 427-437.

[23] H Sherali, A Alameddine. *A new reformulation-linearization technique for bilinear programming problems,* J Global Optim, 1992, 2: 379-410.

[24] H D Sherali, C M Shetty. *A finitely convergent algorithm for bilinear programming problem using polar cuts and disjunctive face cuts,* Math Program, 1980, 19: 379-410.

[25] D Sun. *The strong second order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications,* Math of Oper Res, 2006, 31: 761-776.

[26] D Sun, J Sun, L Zhang. *The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming,* Math Program, 2008, 114: 349-391.

[27] V Visweswaran, C Floudast. *A global optimization algorithm (gop) for certain classes of nonconvex NLPs-ii. application of theory and test problems,* Comput Chem Eng, 1990, 14: 1419-1434.

[28] H Yin, C Ling, L Qi. *Smooth and semismooth Newton methods for constrained approximation and estimation,* Numer Func Anal Opt, 2012, 33: 558-589.

[29] J Zhang, X Chen, X Ding. *Degeneracy removal in cutting plane methods for disjoint bilinear programming,* Optim Lett, 2017, 11: 483-495.

School of Science, Wuhan University of Technology, Wuhan 430070, China.

Email: hesux@whut.edu.cn, wchuanmei@163.com