Noname manuscript No.
(will be inserted by the editor)

Descent Scheme for a Class of Bilevel Programming Problems

Bhuvnesh Khatana · Geetanjali Panda

Received: date / Accepted: date

Abstract In this paper, a class of bilevel programming problems is studied, in which the lower level is a quadratic programming problem, and the upper level problem consists of a nonlinear objective function with coupling constraints. An iterative process is developed to generate a sequence of points, which converges to the solution of this problem. In each iteration of this process, the active set constraints of the lower level problem are modified while decreasing the upper level objective function. Some theoretical results are proven to justify the strategy. Numerical computations are provided to support the methodology.

Keywords Bilevel programming problems · Active set method · Line search method · Armijo condition

Mathematics Subject Classification (2000) 65K10, 90C26, 90C33, 90C53

1 Introduction

The general form of a bilevel programming problem with inequality constraints is stated as

$$(BP) \qquad \begin{aligned} &\underset{x\in\mathbb{R}^n}{\text{“min”}} && F(x,y) \\ &\text{subject to} && G(x,y) \leq 0, \\ & && y \in \Psi(x), \end{aligned}$$

where $\Psi(x) = \{y \mid y \in \underset{y\in\mathbb{R}^m}{\text{argmin}} f(x,y) \text{ subject to } g(x,y) \leq 0\}$, $F : \mathbb{R}^{n+m} \mapsto \mathbb{R}$, $f : \mathbb{R}^{n+m} \mapsto \mathbb{R}$, $G : \mathbb{R}^{n+m} \mapsto \mathbb{R}^p$, $g : \mathbb{R}^{n+m} \mapsto \mathbb{R}^q$.
The optimization problem in the constraint set of the BP is called the lower level problem, which is,

$$(LP_x) \qquad \min_{y\in\mathbb{R}^m} f(x,y) \text{ subject to } g(x,y) \leq 0$$

Here, $\Psi : \mathbb{R}^n \mapsto 2^{\mathbb{R}^m}$ is a point to set mapping, which maps the upper level variable x into the set of optimal solutions of the lower level problem. For each upper level variable x, the set $\Psi(x)$ is called as a rational reaction set. Note that minimization of the function F is obtained with respect to the variable x only, which stays valid when the point to set map Ψ is a singleton set for any x. However, if the lower level problem LP_x has non-unique solutions, then the minimization of F with respect to x may create

Bhuvnesh Khatana
Department of Mathematics,
Indian Institute of Technology Kharagpur,
West Bengal, India
bhuvnesh323001@gmail.com

Geetanjali Panda, Corresponding author
Department of Mathematics,
Indian Institute of Technology Kharagpur,
West Bengal, India
geetanjali@maths.iitkgp.ac.in

ambiguity. Therefore, in the context of BP, we use the term "min" to differentiate this from the word minimization for a general optimization problem.
The set

$$IR = \{(x, y) \mid G(x, y) \leq 0, y \in \Psi(x)\}$$

is known as the "Inducible Region" of BP, and the points in IR are called rational points. IR is known as the feasible set for BP. Denote $X = \{x \in \mathbb{R}^n | G(x, y) \leq 0, y \in \Psi(x)\}$, which is the projection of IR onto $\mathbb{R}^n$.

Optimality conditions for BP are derived by many researchers (see Ye and Zhu ([26], [27]), Dempe and Zemkoho [8], Dempe et al. [7]), Dempe [4]) to justify the existence of its solution. One way to deal with BP is to reformulate this into a single level optimization problem using the optimality condition of the lower level problem, which is known as a mathematical program with equilibrium constraints ($MPEC$) reformulation. As a consequence, the transformed single level problem becomes nonlinear and nonconvex. Under some assumptions, solution methodologies and theoretical developments for $MPEC$ are also applicable to BP. The relation between solutions of BP and respective $MPEC$ may be seen in [6]. Some well-known methods for solving $MPEC$ in this category are regularisation methods ([14] and [23]) and smoothing methods ([10] and [13]). In this paper, we have used an active set strategy on the lower level problem of the bilevel programming problem at a point and constructed a subproblem to obtain suitable direction vectors at that point. KKT optimality conditions on the lower level of this subproblem yield another auxiliary optimization subproblem. To obtain a descent direction at a point, the subproblem is solved corresponding to a suitable active set of lower level problem, which is similar to a disjunctive technique of complementarity constraints of an $MPEC$ framework. Furthermore, to determine an optimal solution for BP, we rely on the nonexistence of descent direction in the contingent cone at a stationary point, where the considered contingent cone corresponds to the respective $MPEC$ formulation of BP.

Various methods like vertex enumeration techniques, KKT approach, and penalty function approach are suggested to solve BP, which are discussed in detail in Bard [1]. Some bilevel programming problems can be solved by heuristic methods including genetic algorithm by Calvete et al. [3], particle swarm optimization methods by Kuo et al. [15] and Han et al. [12], and evolutionary algorithms by Wang et al. [24,25], Sinha et al. [20,22] and [21]. A quadratic bilevel programming problem is a particular case of BP in which the objective functions at both levels are quadratic, and the constraints form a polyhedral structure at both levels. Some traditional methods for quadratic bilevel problems are the branch and bound technique (Bard and Moore [2]) and the enumeration technique (Etoa [9]).

The line search method is another approach to solve BP, which develops a descent sequence that converges to the stationary point of BP. In a general line search method, it is essential to compute the feasible direction at a rational point that decreases the upper level objective function in each iteration. A few research work is available in this direction that is applicable to certain classes of bilevel programming problems. Savard and Gauvin [19] have developed the steepest descent method for a class of BP, in which the upper level problem is free from constraints. At every iteration of this method, a quadratic bilevel problem is solved to obtain the descent direction. Mersha and Dempe [17] used the feasible direction method to solve a nonlinear BP in which the upper level constraints are not dependent on the lower level variable.

In this paper, we deal with the following class of bilevel programming problem, denoted by $CQBP$, in which the upper level problem is not necessarily a quadratic programming problem, and the upper level variable is associated with the coupling constraints. The proposed method is different from the above line search methods. Here, we consider the active constraints at every iteration and obtain the descent direction at the iterating point using a quadratic subproblem with the help of the active constraints. Moreover, we consider the quadratic approximation of the objective function in the subproblem instead of linear approximation in the existing line search schemes.

$$\begin{aligned} (CQBP) \qquad & \min_x && F(x, y) \\ & \text{subject to} && p_i^T x + q_i^T y \leq r_i, \ i \in \Lambda_p, \\ & && \min_y \begin{bmatrix} c_1^T & c_2^T \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \frac{1}{2} \begin{bmatrix} x^T & y^T \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ & && \quad \text{subject to } l_j^T x + m_j^T y \leq n_j, \ j \in \Lambda_q, \end{aligned}$$

where the index set is $\Lambda_k := \{1, 2, ..., k\}$, $p_i \in \mathbb{R}^n,\ q_i \in \mathbb{R}^m,\ r_i \in \mathbb{R}$ for $i \in \Lambda_p,\ l_j \in \mathbb{R}^n,\ m_j \in \mathbb{R}^m,\ n_j \in \mathbb{R}$ for $j \in \Lambda_q$. $c_1 \in \mathbb{R}^n,\ c_2 \in \mathbb{R}^m$, $Q_{21} \in \mathbb{R}^{m\times n}$ $Q_{12} \in \mathbb{R}^{n\times m},\ Q_{11} \in \mathbb{R}^{n\times n},\ Q_{22} \in \mathbb{R}^{m\times m}$.

To focus on the main results, we assume the following restrictive assumptions on $CQBP$ throughout this paper.

(a) The upper level objective function $F : \mathbb{R}^{n+m} \mapsto \mathbb{R}$ is twice continuously differentiable and strictly convex function for both level variables.
(b) The matrix Q_{22} is symmetric and positive definite. (In this case, the lower level problem is convex and has a unique solution corresponding to any feasible point of upper level problem.)
(c) Linear Independence Constraint Qualification ($LICQ$) is satisfied for the lower level problem of $CQBP$, that is, the set $\{m_j \mid l_j^T x + m_j^T y = n_j, j \in \Lambda_q\}$ is linearly independent at a rational point (x, y).

Furthermore, X is a connected subset of $\mathbb{R}^n$, and without loss of generality, we assume that $Q_{21} = Q_{12}^T$.

In this restrictive framework, we use an active set strategy to find a feasible descent direction at a rational point. This direction vector can be related to the contingent cone for the $CQBP$. This paper aims to develop an iterative process to solve $CQBP$. The active set method for a single-level problem is modified to determine a descent direction by solving a quadratic subproblem to maintain a sufficient decrease in the approach. δ-active search technique of [28] is used to ensure the global convergence properties of the method. It is proved that the iteration process converges to a point that satisfies the sufficient optimality condition for the $CQBP$ under the strict convexity assumption of F and boundedness assumption on some level set of F.

This paper is organized into several sections. Section 3 presents the methodology and analysis of our proposed approach, including convergence results. Finally, Section 4 presents a computational experience for the approach and numerical experiments with some test problems.

2 Preliminaries

This section provides some prerequisites related to the general bilevel programming problem, which will be used in the methodology later to solve $CQBP$. The lower level problem LP_x is a parametric optimization problem in the upper level variable x. Consider the Lagrange function for the lower level problem as

$$L_x(y, \lambda) := f(x, y) + \lambda^\top g(x, y),$$

where $\lambda \in \mathbb{R}^q$ is Lagrange multiplier vector. Denote the set of Lagrange multipliers as

$$U_x(y) := \{\lambda \mid \lambda \geq 0,\ \nabla_y f(x, y) + \lambda^T \nabla_y g(x, y) = 0,\ g(x, y) \leq 0,\ \lambda^T g(x, y) = 0\}.$$

Further, let $I_x(y)$ be the active set of lower level problem at any feasible point y, that is,

$$I_x(y) := \{i \in \Lambda_q \mid g_i(x, y) = 0\}.$$

Denote

$$J(\lambda) := \{i \in I_x(y) \mid \lambda_i > 0\}.$$

The definition of the local optimal solution by Dempe [5] takes the following form when coupling constraints are present at the upper level problem.

Definition 2.1 *A point $(x^*, y^*) \in \mathbb{R}^n \times \mathbb{R}^m$ is called a local optimal solution of BP, if $G(x^*, y^*) \leq 0,\ y^* \in \Psi(x^*)$ and there exists an open neighborhood $B_\delta(x^*),\ \delta > 0$, with $F(x, y) \geq F(x^*, y^*)$ for all (x, y) satisfying $x \in B_\delta(x^*), G(x, y) \leq 0,\ y \in \Psi(x)$.*

Notations:
$|,|$ denotes the cardinality of a set and $\|,\|$ denotes the Euclidean norm of a vector.
Suppose $W \subseteq I_x(y) \subseteq \Lambda_q$ for some index sets and let $\lambda \in U_x(y)$. Denote $\lambda_w = (\lambda_j)_{j\in W}$, that is, λ_w is a vector whose components are Lagrange multipliers corresponding to the set W.
$\det(A)$ denotes the determinant of a square matrix A.

3 Methodology

The primal active set method for a general single level quadratic programming problem generates an iterative process that remains feasible at every iteration while decreasing the objective function. At every iteration, the descent direction at that iterating point is determined by solving a subproblem corresponding to a subset of the active set known as the working set. If the iterating point at some iteration minimizes the objective function of the original problem, then the iterative process stops; otherwise, a step is computed to determine the next iterative point and the next working set. Details on the active set method for single level problems can be found in [11, 18]. This paper uses this concept at the lower level of $CQBP$ to determine the descent direction at every iteration. Some important steps of the proposed methodology are stated below.

- At any initial rational point (x^0, y^0), the working set is considered as $W^0 = I_{x^0}(y^0)$.
- At k^{th} iteration, the direction vector $(d^k, w^k) \in \mathbb{R}^n \times \mathbb{R}^m$ is obtained by solving a subproblem in d and w at (x^k, y^k) corresponding to the working set W^k.
- Armijo line search method is used to determine the step length in the direction (d^k, w^k) at rational point (x^k, y^k), and a new point is generated as $(x^{k+1}, y^{k+1}) = (x^k, y^k) + \alpha_k(d^k, w^k)$.
- This process is continued, and a sequence $\{(x^k, y^k)\}$ is generated until the solution of the subproblem yields a zero step.
- Lagrange multipliers of the lower level of $CQBP$ and multipliers at the solution of the subproblem are used together to drop the index from the working set.
- If the solution of the subproblem is a zero vector and the optimality condition is satisfied, then the process is stopped, and the working set is said to be an optimal working set.

Since the inducible feasible region of $CQBP$ is piecewise linear under the convexity assumption of the lower level problem, it is possible to develop a sequence of rational points $\{(x^k, y^k)\}$ moving along the piecewise linear path and converges to a local optimal solution of $CQBP$. Moving along the active set constraints of the lower level problem, we intend to decrease the value of the upper level objective function at every step. To obtain a feasible descent direction, we rely on the active set paradigm for the lower level problem. This process relates to the disjunctive technique for $MPEC$ [16].

Since the constraints of the lower level problem of $CQBP$ are continuous, there exists a neighborhood around a rational point (x, y), so that the inactive constraints of the lower level problem remain unchanged. Hence, obtaining a new point moving along the subset of the active set is possible. We say this set as a working set denoted by W of the $CQBP$, $W \subseteq I_x(y)$, where $I_x(y)$ is the active set for the lower level problem, that is,

$$I_x(y) := \{j \in \Lambda_q \mid l_j^T x + m_j^T y = n_j\}.$$

Let (x, y) be a rational point of $CQBP$ and W be the working set at (x, y). If (x, y) is not an optimal point of $CQBP$, then we can compute the step along the feasible direction (d, w) while decreasing the upper level objective function. For this purpose, consider the following subproblem at the rational point (x, y).

$$P^1_{(x,y)}: \quad \begin{array}{ll} \min\limits_{d} & \begin{bmatrix}\nabla_x F(x,y)^T \ \nabla_y F(x,y)^T\end{bmatrix} \begin{bmatrix} d \\ w \end{bmatrix} + \frac{1}{2} \begin{bmatrix} d \\ w \end{bmatrix}^T \begin{bmatrix} \nabla_{xx} F(x,y) \ \nabla_{xy} F(x,y) \\ \nabla_{yx} F(x,y) \ \nabla_{yy} F(x,y) \end{bmatrix} \begin{bmatrix} d \\ w \end{bmatrix} \\ \text{subject to} & p_i^T d + q_i^T w \le r_i - (p_i^T x + q_i^T y), \ i \in \Lambda_p, \\ & \text{where } w \text{ is the solution of} \\ & \min\limits_{w} \begin{bmatrix} c_1^T \ c_2^T \end{bmatrix} \begin{bmatrix} x + d \\ y + w \end{bmatrix} + \frac{1}{2} \begin{bmatrix} (x+d)^T \ (y+w)^T \end{bmatrix} \begin{bmatrix} Q_{11} \ Q_{12} \\ Q_{21} \ Q_{22} \end{bmatrix} \begin{bmatrix} x + d \\ y + w \end{bmatrix} \\ & \text{subject to } l_j^T d + m_j^T w = 0, \ j \in W. \end{array}$$

The problem $P^1_{(x,y)}$ is comprised of the quadratic approximation of the upper level objective function of $CQBP$ at the rational point (x, y), which is a convex function due to the convexity of F. Lower level constraints of $P^1_{(x,y)}$ are considered in terms of d and w corresponding to the working set constraints in the lower level of $CQBP$.

To maintain the feasibility of $CQBP$ at $(x+d, y+w)$, the lower level inactive constraints can be shifted to the upper level in $P^1_{(x,y)}$. As a result, a new subproblem can be constructed at (x,y) as

$$P^2_{(x,y)}: \quad \begin{array}{ll} \min\limits_{d} & \left[\nabla_x F(x,y)^T \ \nabla_y F(x,y)^T\right] \begin{bmatrix} d \\ w \end{bmatrix} + \frac{1}{2} \begin{bmatrix} d \\ w \end{bmatrix}^T \begin{bmatrix} \nabla_{xx} F(x,y) & \nabla_{xy} F(x,y) \\ \nabla_{yx} F(x,y) & \nabla_{yy} F(x,y) \end{bmatrix} \begin{bmatrix} d \\ w \end{bmatrix} \\ \text{subject to} & p_i^T d + q_i^T w \le r_i - (p_i^T x + q_i^T y), \ i \in \Lambda_p, \\ & l_j^T d + m_j^T w \le n_j - (l_j^T x + m_j^T y), \ j \in \Lambda_q / W, \\ & \min\limits_{w} \left[c_1^T \ c_2^T\right] \begin{bmatrix} x+d \\ y+w \end{bmatrix} + \frac{1}{2} \left[(x+d)^T \ (y+w)^T\right] \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} x+d \\ y+w \end{bmatrix} \\ & \text{subject to } l_j^T d + m_j^T w = 0, \ j \in W. \end{array}$$

The lower level problem of $P^2_{(x,y)}$ is convex and regular, so the lower level problem can be replaced by its KKT optimality conditions, which are

$$c_2 + Q_{12}^T (x+d) + Q_{22}(y+w) + \sum_{j \in W} m_j \mu_j = 0, \tag{1}$$

$$\text{and } l_j^T d + m_j^T w = 0, \ j \in W, \tag{2}$$

where μ_j are the dual variables associated with constraints $l_j^T d + m_j^T w = 0$, for $j \in W$. If the active set strategy is applied at $(x+d, y+w)$ for the lower level problem of $CQBP$ then one may observe that (1) and (2) are satisfied in case $\mu_j \ge 0, \ j \in W$. This system is feasible for the working set $J(\lambda) \subseteq W \subseteq I_x(y)$. Therefore we may restrict $P^2_{(x,y)}$ with $\mu_j \ge 0, \ j \in W$ and construct the following subproblem at (x,y), given the working set W, as

$QP_{(x,y)}(W)$:

$$\min_{(d,w,\mu) \in \Omega_{(x,y)}(W)} \left[\nabla_x F(x,y)^T \ \nabla_y F(x,y)^T\right] \begin{bmatrix} d \\ w \end{bmatrix} + \frac{1}{2} \begin{bmatrix} d \\ w \end{bmatrix}^T \begin{bmatrix} \nabla_{xx} F(x,y) & \nabla_{xy} F(x,y) \\ \nabla_{yx} F(x,y) & \nabla_{yy} F(x,y) \end{bmatrix} \begin{bmatrix} d \\ w \end{bmatrix},$$

where

$$\Omega_{(x,y)}(W) := \left\{ (d,w,\mu) \in \mathbb{R}^{(m+n+n_w)} : \begin{array}{l} p_i^T d + q_i^T w \le r_i - (p_i^T x + q_i^T y), \ i \in \Lambda_p, \\ c_2 + Q_{12}^T (x+d) + Q_{22}(y+w) + \sum\limits_{j \in W} m_j \mu_j = 0 \\ l_j^T d + m_j^T w = 0 \ , \ j \in W, \\ l_j^T d + m_j^T w \le n_j - (l_j^T x + m_j^T y), \ j \in \Lambda_q / W, \\ \mu_j \ge 0, \ j \in W \end{array} \right\}.$$

The subproblem $QP_{(x,y)}(W)$ has a unique global minimum since F is assumed as a strict convex function. Moreover, $LICQ$ holds at the solution of the lower level problem of $CQBP$. Hence the corresponding dual vector λ is unique, and $J(\lambda)$ is a fixed index set. The feasible directions at a rational point (x,y) can be obtained by projecting the set $\bigcup\limits_{J(\lambda) \subseteq W \subseteq I_x(y)} \Omega_{(x,y)}(W)$ onto $\mathbb{R}^n \times \mathbb{R}^m$. The contingent cone for $CQBP$ in case of coupling constraints (see Section (5.3) of [5]) is

$$\hat{C}(x,y) = \{(d,w) \mid w = y'(x,d), \ p_i^T d + q_i^T w \le 0, \ i \in \{i : p_i^T x + q_i^T y = r_i\}\},$$

where $y'(x,d)$ denotes the directional derivative of the optimal solution $y(x)$ of the lower level problem of $CQBP$, which can be determined using System (3) of Lemma 1 below. The set $\Omega_{(x,y)}(W)$ can be related to this system through the transformation $\mu_j = \lambda_j + \gamma_j$, where $j \in W$ and γ is defined in the lemma below. In that case, the contingent cone is equivalent to the set of feasible directions.

Lemma 3.1 *Let (x,y) be a rational point of $CQBP$. Then, $w = y'(x,d)$ satisfies the following system at d for some index set W satisfying $J(\lambda) \subseteq W \subseteq I_x(y)$ and $\gamma \in \mathbb{R}^{|I_x(y)|}$, where $|I_x(y)|$ denotes the cardinality of the index set $I_x(y)$.*

$$\left. \begin{array}{l} Q_{12}^T d + Q_{22} w + \sum\limits_{j \in I_x(y)} m_j \gamma_j = 0 \\ l_j^T d + m_j^T w = 0, \ j \in W \\ l_j^T d + m_j^T w \le 0, \ j \in I_x(y)/W \\ \gamma_j \ge 0, \ j \in W/J(\lambda), \quad \gamma_j = 0, \ j \notin W \end{array} \right\} \tag{3}$$

Proof Using Theorem 4.11 of Dempe [5] for general bilevel programming problem, one can conclude that the directional derivative of $y(x)$ of $CQBP$ along the direction d coincides with the unique optimal solution of the convex quadratic programming problem

$$\begin{array}{lll} \min\limits_{w} & \frac{1}{2}w^T Q_{22} w + w^T Q_{12} d & \\ \text{subject to} & l_j^T d + m_j^T w = 0, & i \in J(\lambda) \\ & l_j^T d + m_j^T w \leq 0, & i \in I_x(y)/J(\lambda) \end{array} \tag{4}$$

Hence, using the active set strategy for the active set W, System (3) can be obtained, where w is the optimal solution and γ is the unique Lagrange multiplier vector of Problem (4) since $LICQ$ is considered for the lower level problem of $CQBP$. Hence, the result follows. □

The following result is motivated by Theorem 5.5 of [5].

Theorem 3.1 *(Sufficient optimality condition of $CQBP$) Suppose (d_W, w_W, μ_W) be the optimal solution of following problem for some W satisfying $J(\lambda) \subseteq W \subseteq I_x(y)$.*

$$\min_{d,w,\mu} \nabla_x F(x,y)^T d + \nabla_y F(x,y)^T w$$
$$\textit{subject to } (d,w,\mu) \in \Omega_{(x,y)}(W) \textit{ and } \|d\| = 1,$$

and if

$$\min\{\nabla_x F(x,y)^T d_W + \nabla_y F(x,y)^T w_W \mid J(\lambda) \subseteq W \subseteq I_x(y)\} \geq 0, \tag{5}$$

then (x,y) is a strict local minimum of $CQBP$.

Proof If (5) holds and F is a strict convex function, for $\alpha > 0$

$$F(x+\alpha d, y+\alpha w) - F(x,y) > \alpha(\nabla_x F(x,y)^T d + \nabla_y F(x,y)^T w) \geq 0$$

for all $(d,w,\mu) \in \Omega_{(x,y)}(W)$ and $\|d\| = 1$.
Therefore for some $\alpha_0 > 0$ $F(x,y) < F(x+\alpha d, y+\alpha w)$ $\forall \alpha \in (0,\alpha_0]$ for all $(d,w,\mu) \in \Omega_{(x,y)}(W)$, $\|d\| = 1$.
From Lemma 3.1, it can be concluded that $\hat{C}(x,y)$ is equal to the projection of $\bigcup\limits_{J(\lambda)\subseteq W \subseteq I_x(y)} \Omega_{(x,y)}(W)$ onto $\mathbb{R}^n \times \mathbb{R}^m$. Therefore $F(x,y) < F(x+\alpha d, y+\alpha w)$ $\forall \alpha \in (0,\alpha_0]$ for all $(d,w) \in \hat{C}(x,y)$, $\|d\| = 1$. Hence (x,y) is a strict local minimum. □

3.1 General Outline of the Method

Note that $(0, 0, \lambda_{w^k})$ is a feasible point of problem $QP_{(x^k,y^k)}(W^k)$. Later it is proved that if (d^k, w^k, μ^k) is the optimal solution of $QP_{(x^k,y^k)}(W^k)$ and $(d^k, w^k) \neq (0,0)$ then (d^k, w^k) is a feasible descent direction of $CQBP$ at (x^k, y^k). In that case, the iterating point (x^k, y^k) can be updated till it satisfies the optimality criteria. Otherwise, either the current iterating point satisfies the optimality criteria or a new subproblem $QP_{(x^k,y^k)}(W^k)$ can be constructed by modifying the working set W of the lower level problem of $CQBP$ to get nonzero (d, w). The working set can be updated in each iteration till the optimal working set of the $CQBP$ is achieved. In this process, we might need to add and remove constraints from the working set of the lower level problem in every iteration and modify $QP_{(x^k,y^k)}(W^k)$.

Consider the set $S(W, J(\lambda))$ for $J(\lambda) \subseteq W$ as

$$S(W,J(\lambda)) := \left\{ (x,y) \in \mathbb{R}^{(m+n)} \;:\; \begin{array}{l} p_i^T x + q_i^T y \leq r_i,\ i \in \Lambda_p, \\ c_2 + Q_{12}^T x + Q_{22} y + \sum\limits_{j \in W} m_j \lambda_j = 0, \\ l_j^T x + m_j^T y = n_j\ ,\ j \in W, \\ l_j^T x + m_j^T y \leq n_j,\ j \in \Lambda_q/W, \\ \lambda_j \geq 0,\ j \in J(\lambda),\ \lambda_j = 0,\ j \in W/J(\lambda) \end{array} \right\}.$$

The sequence $\{(x^k, y^k)\}$ may converge to a stationary point of F, restricted over $S(W, J(\lambda))$ corresponding to the working set W, instead of converging to a stationary point of $CQBP$. To address this issue, we incorporate the $\delta-$ active search technique concepts, which is implemented in Step 1 of Algorithm 1. This technique is in the light of the active search method in [28].

Consider the $\delta'-$active set as

$$I_k(\delta') := \{j \in \Lambda_q \mid -\delta' \leq l_j^T x^k + m_j^T y^k - n_j \leq 0\}$$
$$\text{and } J_k(\delta') := \{j \in I_{x^k}(y^k) \mid \lambda_j^k > \delta'\} \text{ for some } \delta' > 0.$$

$J_k(\delta') \subseteq I_{x^k}(y^k)$ and $I_{x^k}(y^k) \subseteq I_k(\delta')$. Hence $J_k(\delta') \subseteq I_k(\delta')$. Let the projection of (x^k, y^k) onto $S(I_k(\delta'), J_k(\delta'))$ be denoted by $(\bar{x}, \bar{y})$. Let $(\hat{d}, \hat{w}, \hat{\mu})$ be the optimal solution of the following least mean square problem at (x^k, y^k) when $I_{x^k}(y^k) \neq I_k(\delta')$ or $J(\lambda^k) \neq J_k(\delta')$, and $S(I_k(\delta'), J_k(\delta')) \neq \emptyset$.

$$\begin{aligned}
\underset{d,w,\mu}{\text{minimize}} \quad & ||d||^2 + ||w||^2 \\
\text{subject to} \quad & p_i^T(x^k + d) + q_i^T(y^k + w) \leq r_i, \quad i \in \Lambda_p, \\
& c_2 + Q_{12}^T(x^k + d) + Q_{22}(y^k + w) + \sum_{j \in I_k(\delta')} m_j(\lambda_j^k + \mu_j) = 0, \\
& l_j^T(x^k + d) + m_j^T(y^k + w) = n_j, \quad j \in I_k(\delta'), \\
& l_j^T(x^k + d) + m_j^T(y^k + w) \leq n_j, \quad j \in \Lambda_q / I_k(\delta'), \\
& \lambda_j^k + \mu_j \geq 0,\ j \in J_k(\delta'),\ \lambda_j^k + \mu_j = 0,\ j \in I_k(\delta')/J_k(\delta').
\end{aligned}$$

Denote

$$(\bar{x}, \bar{y}) := (x^k + \hat{d}, y^k + \hat{w}) \quad \bar{\lambda} = \lambda^k + \hat{\mu}. \tag{6}$$

$\bar{\lambda}$ is uniquely determined since $LICQ$ is satisfied for the lower level problem.

Note that $(\bar{x}, \bar{y}) \in S(I_k(\delta'), J_k(\delta'))$ and $I_{\bar{x}}(\bar{y}) \supseteq I_k(\delta')$ and $J(\bar{\lambda}) \subseteq J_k(\delta')$. In δ active search technique, we search for the possible working set within a certain distance from the current iterative point. That is, when $I_{x^k}(y^k) \neq I_k(\delta')$ or $J(\lambda^k) \neq J(\delta')$, we project the (x^k, y^k) on $S(I_k(\delta'), J_k(\delta'))(\neq \emptyset)$ and obtain an intermediate point $(\bar{x}, \bar{y})$ to facilitate a change in the working set. However, to ensure the global convergence of the approach, we update to the next point if the condition

$$||\nabla_x F(x^k, y^k) d^k + \nabla_y F(x^k, y^k) w^k|| > \epsilon$$

is satisfied for some $\epsilon > 0$ to maintain the sufficient decrease in the approach. Otherwise, we modify the working set to obtain a feasible descent direction set to decrease the function F such that $||\nabla_x F(x^k, y^k) d^k + \nabla_y F(x^k, y^k) w^k|| > \epsilon$. At every iteration, the quadratic subproblem $QP_{(x^k,y^k)}(W^k)$ is solved. Let the solution of this subproblem be (d^k, w^k, μ^k).

Case 1:
$||\nabla_x F(x^k, y^k) d^k + \nabla_y F(x^k, y^k) w^k|| > \epsilon$
In this case, the next iterating point is obtained by the updating formula

$$(x^{k+1}, y^{k+1}) := (x^k + \alpha_k d^k, y^k + \alpha_k w^k).$$

The Armijo step length procedure can compute the step length α_k. At this new iterating point, a new working set W^{k+1} can be obtained by adding the constraints from the lower level problem of $CQBP$, which are active at (x^{k+1}, y^{k+1}). It is determined by the set

$$T^k = \{j \in \Lambda_q / W^k | \, l_j^T d^k + m_j^T w^k = n_i - (l_j^T x^k + m_j^T y^k)\}.$$

Then working set is updated as $W^{k+1} = W^k \cup T^k$ and the new subproblem $QP_{(x^{k+1},y^{k+1})}(W^{k+1})$ can be taken care for next iteration.

Case 2: $||\nabla_x F(x^k, y^k) d^k + \nabla_y F(x^k, y^k) w^k|| \leq \epsilon$
In this case, we modify the working set to ensure the condition of Case 1. For this, we use the idea of active set methods for single level problems. At a rational point (x^k, y^k), denote

$$v(x^k, y^k) := \begin{bmatrix} \nabla_x F(x^k, y^k) \\ \nabla_y F(x^k, y^k) \\ 0 \end{bmatrix},$$

where $0 \in \mathbb{R}^{n_w}$ is the null vector and $n_w = |W|$. Collect the constraints from $\Omega_{(x^k,y^k)}(W^k)$ which are active at the solution (d^k, w^k, μ^k). Let A^k be the coefficient matrix of those active constraints. Suppose a submatrix A_{W^k} be derived from A^k in such a way that

- A_{W^k} must include those rows of A^k, which correspond to the equality constraints of $QP_{(x^k,y^k)}(W^k)$,
- A_{W^k} has full row rank, and
- $A^T_{W^k} z = -v(x^k, y^k)$ is consistent.

Let $z(= z_s)$ be the unique solution of the system $A^T_{W^k} z = -v(x^k, y^k)$. The working set W^k can be modified by a suitable process to get suitable direction vector (d^k, w^k). We select an index corresponding to the minimum value of z_s for the index i_s in $W^k/J(\lambda^k)$ and a new subproblem $QP_{(x^k,y^k)}(W^k/i_s)$ is constructed. The process can be repeated until $||\nabla_x F(x^k, y^k)d^k + \nabla_y F(x^k, y^k)w^k|| > \epsilon$. Here, the following situations arise.

(a) The system $A^T_{W^k} z = -v(x^k, y^k)$ is not consistent. Then, we reduce the value of ϵ in the process.
(b) The condition of Case 1, $||\nabla_x F(x^k, y^k)d^k + \nabla_y F(x^k, y^k)w^k|| > \epsilon$ can not be obtained for all the choices of W^k satisfying $J(\lambda^k) \subseteq W^k \subseteq I_{x^k}(y^k)$, then we select (d^k, w^k) for which $||\nabla_x F(x^k, y^k)d^k + \nabla_y F(x^k, y^k)w^k||$ is maximum and the next point has to be updated.

We use the following rule to terminate the algorithm. An iterative point (x^k, y^k) is considered as a stopping point of the method if $(d^k, w^k) = (0, 0)$ and any of the following conditions are satisfied.

(c) If $z_s \geq 0$ corresponding to the working set constraints $W^k/J(\lambda^k)$ and the inequality constraints.
(d) If the solution of $QP_{(x^k,y^k)}(W^k)$ is $(0, 0, \lambda_{w^k})$ for every choice of W^k satisfying $J(\lambda^k) \subseteq W^k \subseteq I_{x^k}(y^k)$.

The stopping point satisfies the sufficient optimality criteria of $CQBP$, which is proved in Theorem 3.5 in this section.

3.2 Construction of A_{W^k} at (x^k, y^k)

Note that A_{W^k} is obtained from A so that A_{W^k} has full row rank. In addition to this, A_{W^k} is not necessarily unique. If the obtained matrix A_{W^k} has a maximum number of rows maintaining full row rank, then the system $A^T_{W^k} z = -v(x^k, y^k)$ is consistent. Denote the following matrices:

$$L_w := \{l_j\}_{j \in W^k} \in \mathbb{R}^{n \times n_w}, \ M_w := \{m_j\}_{j \in W^k} \in \mathbb{R}^{m \times n_w}, \text{ where } n_w = |W|.$$

Let $K^k \subseteq \Lambda_p$ such that $\{p_i \mid i \in K^k\}$ is a linearly independent set.

$$P_k := \{p_i\}_{i \in K^k} \in \mathbb{R}^{n \times m_k}, \ Q_k := \{q_i\}_{i \in K^k} \in \mathbb{R}^{m \times m_k}, \text{ where } m_k = |K^k|.$$

Lemma 3.2 *The matrix*

$$\begin{bmatrix} Q_{12}^T & Q_{22} & M_w \\ L_w^T & M_w^T & 0 \\ P_k^T & Q_k^T & 0 \end{bmatrix} \tag{7}$$

has full row rank.

Proof Consider the submatrix $B = \begin{bmatrix} Q_{22} & M_w \\ M_w^T & 0 \end{bmatrix}$. Using Schur complement theorem on the matrix $B \in \mathbb{R}^{(m+n_w)\times(m+n_w)}$, we have

$$\det(B) = \det(Q_{22}) \det(M_w^T Q_{22}^{-1} M_w).$$

Since Q_{22} is positive defnite matrix, $\det(Q_{22}) \neq 0$. Furthermore, M_w is a full column rank matrix, then $\det(M_w^T Q_{22}^{-1} M_w) \neq 0$. Hence, B is a full rank matrix. P_k^T is a full-row rank matrix. Therefore, the matrix in (7) is a full row rank matrix. Hence, the result follows. □

The selection of the matrix A_{W^k} at (x^k, y^k) can be challenging. Since the matrix in (7) has full row rank in a particular case, the effort to obtain A_{W^k} can be reduced. In that case the matrix A_{W^k} can be constructed by adding some rows from the active set matrix A^k to the matrix in (7) so that A_{W^k} will be full row rank. Otherwise, A_{W^k} can be obtained from the matrix A^k by eliminating some rows so that the resulting matrix is a full-row rank matrix.

Let $z(= z_s)$, where z_s is component of the vector z corresponding s^{th} column vector of matrix $A^T_{W^k}$ and i_s be the index of the constraint in the working set, which is associated with the s^{th} column vector of the matrix $A^T_{W^k}$. Suppose

$$z_{s^*} := \min\{z_s \mid \lambda^k_{i_s} = 0, i_s \in W^k\}, \tag{8}$$

and minimum of the set occurs at the index s^* for which the corresponding index in the working set is i_{s^*}.

3.3 Determination of Descent Direction at a Nonstopping Point

Case (2c) and Case (2d) determine the method's stopping point. If the solution of the subproblem is nonzero at (x^k, y^k) then Corollary 3.1 proves that the direction vector is descent to F. Otherwise, If $(0, 0, \lambda_{w^k})$ is optimal solution of $QP_{(x^k,y^k)}(W^k)$ at some iterate (x^k, y^k) for some appropriate W^k. Further, if (x^k, y^k) is not stopping point then there is at least one index s with $z_s < 0$ corresponding to the inequality constraints or working set constraint in $W^k/J(\lambda^k)$. In this case, the direction vector can be described in two cases.

Case (i): Suppose $A_{W^k} = A^k$, where A^k is the active set matrix at $(0, 0, \lambda_{w^k})$ then A_{W^k} is also active at $(0, 0, \lambda_{w^k})$. Hence

$$A_{W^k} \begin{bmatrix} 0 \\ 0 \\ \lambda_{w^k} \end{bmatrix} = \begin{bmatrix} -(c_2 + (Q^T_{12} x^k + Q_{22} y^k)) \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{9}$$

Suppose the case where $z_s < 0$ corresponds to $W^k \setminus J(\lambda^k)$ only. Denote a vector $a \in \mathbb{R}^{n_w}$ as

$$a_i = \begin{cases} 0, & i \neq i_{s^*}, \\ < 0, & i = i_{s^*}. \end{cases}$$

A_{W^k} is a full-row rank matrix. Then there is some (d', w', μ'), which satisfies the following system

$$A_{W^k} \begin{bmatrix} d' \\ w' \\ \mu' \end{bmatrix} = \begin{bmatrix} 0 \\ a \\ 0 \\ 0 \end{bmatrix}. \tag{10}$$

This implies

$$v(x^k, y^k)^T \begin{bmatrix} d' \\ w' \\ \mu' \end{bmatrix} = -z^T A_{W^k}. \begin{bmatrix} d' \\ w' \\ \mu' \end{bmatrix} = -z^T \begin{bmatrix} 0 \\ a \\ 0 \\ 0 \end{bmatrix} = -z_{s^*} a_{i_{s^*}} < 0, \tag{11}$$

where s^* is the index obtained from (8) and $z_{s^*} < 0$. Adding (9) and (10), then using the property of vector a, we get

$$Q^T_{12} d' + Q^T_{22} w' + M_w(\lambda_{w^k} + \mu') = c_2 + Q^T_{12} x + Q^T_{22} y, \tag{12}$$

$$l^T_j d' + m^T_j w' < 0 \; j \in i_{s^*}, \tag{13}$$

$$l^T_j d' + m^T_j w' = 0, \; j \in W^k/i_{s^*}, \tag{14}$$

$$p^T_i d + q^T_i w \leq 0, \; i \in K^k, \tag{15}$$

$$\mu_j = 0, j \in W^k/J(\lambda^k). \tag{16}$$

Substituting the value of $v(x^k, y^k)$ in (11), we have

$$\nabla_x F(x^k, y^k)^T d' + \nabla_y F(x^k, y^k)^T w' < 0. \tag{17}$$

This guarantees that $(d', w') \neq (0,0)$. From (12)-(16) and (17), it is confirmed that there exists a descent direction (d', w'), which is a feasible solution of $QP_{(x^k,y^k)}(W^k/i_{s^*})$. Hence if $(d^k, w^k) = (0,0)$ at k^{th} iteration then W^k can be replaced by W^k/i_{s^*} so that $QP_{(x^k,y^k)}(W^k/i_{s^*})$ has a nonzero solution which is a descent direction at (x^k, y^k). In fact, $z_s < 0$ can not correspond to inequality constraints for this case. If $z_s < 0$ corresponds to inequality constraints, then using the above procedure, one can find a feasible descent direction. That can not be possible because at (x^k, y^k), $(0, 0, \lambda_w^k)$ is the optimal solution of $QP_{(x^k,y^k)}(W^k)$.

Case (ii): Suppose $A_{W^k} \neq A^k$. From Case (2.d), there exits (d'', w'', μ'') with $(d'', w'') \neq (0,0)$, which is feasible to $\Omega_{(x^k,y^k)}(W^k)$ for some W^k. It can be seen that (d'', w'', μ'') satisfies the system

$A_{W^k} \begin{bmatrix} d'' \\ w'' \\ \mu'' \end{bmatrix} = \begin{bmatrix} 0 \\ a \\ b \\ c \end{bmatrix}$, for some $a, b, c \leq 0$. Then, (d'', w'', μ'') is descent since

$$v(x^k, y^k)^T \begin{bmatrix} d'' \\ w'' \\ \mu'' \end{bmatrix} = -z^T A_{W^k} \begin{bmatrix} d'' \\ w'' \\ \mu'' \end{bmatrix} = -z^T \begin{bmatrix} 0 \\ a \\ b \\ c \end{bmatrix} < 0 \tag{18}$$

can be obtained proceeding as in Case I for some a, b and c.

The above discussion guarantees that a nonzero decent vector can be obtained by modifying the working set in the subproblem.

The above aspects are summarised in the following algorithm for the descent scheme for $CQBP$.

Algorithm 1: Descent scheme for $CQBP$

Step 0: Select initial rational point (x^0, y^0), initial working set $W^0 = I_{x^0}(y^0)$, and parameters $0 < \delta < \beta < 1,\ \sigma_1, \sigma_2 \in (0,1),\ \epsilon > 0$.
Step 1: Set $\delta' = \delta_0$;
if $I_k(\delta') = I_{x^k}(y^k)$ *and* $J_k(\delta') = J(\lambda^k)$ **then**
  $W^k \leftarrow I_{x^k}(y^k)$, go to Step 2.
else
  if $S(I_k(\delta'), J_k(\delta')) \neq \emptyset$ *and* $F(\bar{x}, \bar{y}) < F(x^k, y^k)$ **then**
    $(x^k, y^k) \leftarrow (\bar{x}, \bar{y})$ (from (6)), $W^k \leftarrow I_{\bar{x}}(\bar{y})$, and Go to Step 2.
  else
    $\delta' = \sigma_1 \max\{n_j - l_j^T x^k - m_j^T y^k, i \in I_k(\delta'),\ \lambda_j^k,\ j \in I_k(\delta') \setminus J_k(\delta')\}$
    and start over from Step 1.
  end
end
Step 2: Solve $QP_{(x^k,y^k)}(W^k)$ and obtain the solution (d^k, w^k, μ^k).
Step 3: (Nested loops)
if $\|\nabla_x F(x^k, y^k) d^k + \nabla_y F(x^k, y^k) w^k\| \leq \epsilon$ **then**
  Compute the matrix A_{W^k}.
  if $A_{W^k}^T z \neq -v(x^k, y^k)$ **then**
    $\epsilon = \sigma_2 \epsilon$;
    Update to the next point (else part of Step 3).
  else
    if (x^k, y^k) *is a stopping point* **then**
      Terminate with the optimal solution (x^k, y^k).
    else
      $Z^k = \text{argmin}\{z_s \mid \lambda_{i_s}^k = 0, i_s \in W^k\}$
      if $W^k \neq \emptyset$ **then**
        Update $W^k \leftarrow W^k / \{i_{s^*}\}$;
        Go to step 2.
      else
        Select any W^k which has not been tried before;
        Otherwise if all W^k are tried, set $\epsilon = \sigma_2 \epsilon$,
        and proceed as Case 2b.
      end
    end
  end
else
  Set $(x^{k+1}, y^{k+1}) = (x^k, y^k) + \alpha_k (d^k, w^k)$,
  Compute $\alpha_k > 0$ such that $\alpha_k = \beta^j,\ j \in \{0, 1, 2...\}$ is the smallest integer satisfying Armijo condition
  $F(x^{k+1}, y^{k+1}) \leq F(x^k, y^k) + \alpha_k \delta (\nabla_x F(x^k, y^k)^T d^k + \nabla_y F(x^k, y^k)^T w^k)$.
  Obtain $T^k = \{j \in \Lambda_q / W^k |\ l_j^T (x^k + d^k) + m_j^T (y^k + w^k) = n_j\}$;
  Set $W^{k+1} \leftarrow W^k \cup T^k$;
  $k = k + 1$;
  Go to step 1.
end

The following theorem justifies that every point of the iteration (x^k, y^k) in Algorithm 1 is a rational point for a given initial rational point (x^0, y^0). It also provides the computation of Lagrange multipliers of the lower level problem at the next rational point.

Theorem 3.2 *If (x^k, y^k) is a rational point of $CQBP$, then (x^{k+1}, y^{k+1}), which is obtained from Algorithm 1 is also a rational point.*

Proof Since (x^{k+1}, y^{k+1}) satisfies the upper level constraints, it remains to show that $y^{k+1} \in \Psi(x^{k+1})$. The iterative points are updated in Step 1 and Step 3 of Algorithm 1.
Suppose from Step 1, $(x^k, y^k) \leftarrow (\bar{x}, \bar{y})$. Clearly $\bar{y} \in \Psi(\bar{x})$. Otherwise, the next point is obtained by solving $QP_{(x^k,y^k)}(W^k)$ for some W^k satisfying $J(\lambda^k) \subseteq W^k \subseteq I_{x^k}(y^k)$. If (d^k, w^k, μ^k) is solution of $QP_{(x^k,y^k)}(W^k)$ then we show that $y^{k+1} \in \Psi(x^{k+1})$.

Using the transformation $\mu_j = \lambda_j + \gamma_j,\ j \in W^k$ in the constraint set $\Omega_{(x^k,y^k)}(W^k)$, (d^k, w^k, γ^k) satisfies the following system:

$$\left.\begin{aligned}
&Q_{12}^T d^k + Q_{22} w^k + \sum_{j \in W^k} m_j \gamma_j^k = 0,\\
&l_j^T d^k + m_j^T w^k = 0,\ j \in W^k,\\
&l_j^T (x^k + d^k) + m_j^T (y^k + w^k) \le n_j,\ j \in \Lambda_q / W^k,\\
&\lambda_j^k + \gamma_j^k \ge 0,\ j \in W^k.
\end{aligned}\right\} \tag{19}$$

Using the active set strategy at (x^k, y^k) for the lower level problem,

$$\left.\begin{aligned}
&c_2 + Q_{12}^T x^k + Q_{22} y^k + \sum_{j \in I_{x^k}(y^k)} m_j \lambda_j^k = 0,\\
&l_j^T x^k + m_j^T y^k = n_j,\ j \in I_{x^k}(y^k),\\
&l_j^T x^k + m_j^T y^k < n_j,\ j \in \Lambda_q / I_{x^k}(y^k),\\
&\lambda_j^k = 0, j \in \Lambda_q / I_{x^k}(y^k),\ \lambda_j^k \ge 0, j \in I_{x^k}(y^k).
\end{aligned}\right\} \tag{20}$$

Using (19) and (20), it can be easily verified that $(x^k + \alpha_k d^k, y^k + \alpha_k w^k)$ satisfies the following conditions for each $\alpha_k \in [0, 1]$.

$$\begin{aligned}
&c_2 + Q_{12}^T (x^k + \alpha_k d^k) + Q_{22} (y^k + \alpha_k w^k) + \sum_{j \in W^k} m_j (\lambda_j^k + \alpha_k \gamma_j^k) = 0,\\
&l_j^T (x^k + \alpha_k d^k) + m_j^T (y^k + \alpha_k w^k) = 0,\ j \in W^k,\\
&l_j^T (x^k + \alpha_k d^k) + m_j^T (y^k + \alpha_k w^k) \le n_j,\ j \in \Lambda_q / W^k,\\
&\lambda_j^k + \alpha_k \gamma_j^k \ge 0,\ j \in W^k.
\end{aligned}$$

Hence (x^{k+1}, y^{k+1}) is a rational point for each $\alpha_k \in [0, 1]$. □

Corollary 3.1 *If (d^k, w^k, μ^k) is the solution of the problem $QP_{(x^k,y^k)}(W^k)$ with $(d^k, w^k) \ne 0$ then there exists $\alpha^* > 0$ such that*

$$F(x^k + \alpha d^k, y^k + \alpha w^k) < F(x^k, y^k),\ \forall\ \alpha \in (0, \alpha^*].$$

Proof $(0, 0, \lambda_{w^k})$ is a feasible point of $QP_{(x^k,y^k)}(W^k)$ with objective value zero. (d^k, w^k, μ^k) is the unique optimal solution of $QP_{(x^k,y^k)}(W^k)$ with $(d^k, w^k) \ne (0, 0)$. Hence

$$\begin{bmatrix} \nabla_x F(x^k, y^k)^T & \nabla_y F(x^k, y^k)^T \end{bmatrix} \begin{bmatrix} d^k \\ w^k \end{bmatrix} + \frac{1}{2} \begin{bmatrix} d^k \\ w^k \end{bmatrix}^T \begin{bmatrix} \nabla_{xx} F(x^k, y^k) & \nabla_{xy} F(x^k, y^k) \\ \nabla_{yx} F(x^k, y^k) & \nabla_{yy} F(x^k, y^k) \end{bmatrix} \begin{bmatrix} d^k \\ w^k \end{bmatrix} < 0.$$

The left side of the above inequality is positive as F is a strictly convex function. Therefore

$$\nabla_x F(x^k, y^k)^T d^k + \nabla_y F(x^k, y^k)^T w^k < 0.$$

Hence the result holds. □

Theorem 3.3 *If F is a strictly convex quadratic function, then the step length in Algorithm 1 is $\alpha_k = 1$ in each iteration.*

Proof Since F is a quadratic function, at the iterative point (x^k, y^k),

$$F(x^k + d^k, y^k + w^k) = F(x^k, y^k) + \left[\nabla_x F(x^k, y^k)^T \ \nabla_y F(x^k, y^k)^T\right] \begin{bmatrix} d^k \\ w^k \end{bmatrix} + \frac{1}{2} \begin{bmatrix} d^k \\ w^k \end{bmatrix}^T \begin{bmatrix} \nabla_{xx} F(x^k, y^k) & \nabla_{xy} F(x^k, y^k) \\ \nabla_{yx} F(x^k, y^k) & \nabla_{yy} F(x^k, y^k) \end{bmatrix} \begin{bmatrix} d^k \\ w^k \end{bmatrix}. \quad (21)$$

Further, F is a strictly convex function and $(0, 0, \lambda_{w^k})$ is feasible solution of the problem $QP_{(x^k,y^k)}(W^k)$. If (d^k, w^k, μ^k) is the solution of this problem with $(d^k, w^k) \neq 0$ then

$$\left[\nabla_x F(x^k, y^k)^T \ \nabla_y F(x^k, y^k)^T\right] \begin{bmatrix} d^k \\ w^k \end{bmatrix} + \frac{1}{2} \begin{bmatrix} d^k \\ w^k \end{bmatrix}^T \begin{bmatrix} \nabla_{xx} F(x^k, y^k) & \nabla_{xy} F(x^k, y^k) \\ \nabla_{yx} F(x^k, y^k) & \nabla_{yy} F(x^k, y^k) \end{bmatrix} \begin{bmatrix} d^k \\ w^k \end{bmatrix} < 0.$$

From (21), $F(x^k + d^k, y^k + w^k) < F(x^k, y^k)$ and there exist δ such that

$$F(x^k + d^k, y^k + w^k) \leq F(x^k, y^k) + \delta(\nabla_x F(x^k, y^k)^T d^k + \nabla_y F(x^k, y^k)^T w^k).$$

This inequality implies that $\alpha_k = 1$ satisfies the Armijio condition. □

Theorem (3.3) proves that $\alpha_k = 1$ for every k if the upper level objective function is quadratic. In that case, F attains its global minimum in one step over the feasible set $S(W^k, J(\lambda^k))$, provided that W^k corresponds to a nonzero step. The working set is modified as $W^k \cup T^k$ in the next iteration. Thus, the working set changes in each iteration. The algorithm never returns to the same working set since the objective value decreases at every iteration. In addition, the algorithm seeks the optimal working set, which will be obtained after a finite number of iterations since the working set has a finite number of choices.

The theorem below justifies the convergence of sequence generated by Algorithm 1. To justify the convergence of the sequence $\{(x^k, y^k)\}$ generated by Algorithm 1, we assume that this sequence is bounded in a level set of F. In the limiting case, we denote the final working set of this iteration as W^* and the corresponding vector of Lagrange multipliers as λ^*. Further, μ^k becomes λ_{w^*}, whose components are the components of λ^*, corresponding to the working set W^* satisfying $J(\lambda^*) \subseteq W^* \subseteq I_{x^*}(y^*)$. Before proceeding to the main result, first we require the following lemma.

Lemma 3.3 *Suppose a subsequence of the infinite sequence $\{(x^k, y^k, \lambda^k)\}$ generated by Algorithm 1 converges to an accumulation point (x^*, y^*, λ^*). For each sufficiently large k, it follows that $I_k(\delta^{t_k}) = I_{x^*}(y^*)$, $J_k(\delta^{t_k}) = J(\lambda^*)$, and $S_k(\delta^{t_k}) = S(I_{x^*}(y^*), J(\lambda^*))$, where δ^{t_k} is used in the inner iteration of Step 1 at k^{th} iterate.*

Proof The proof follows from Lemma 3.2 of [28] by considering the index set $I(z^k)$ of [28] as $I_{x^k}(y^k) \cup (I_{x^k}(y^k) \setminus J(\lambda^k))$. □

Theorem 3.4 *Suppose (x^*, y^*, λ^*) is an accumulation point of the sequence $\{(x^k, y^k, \lambda^k)\}$ generated by Algorithm1. Then there must be subsequence K such that $\lim_{k\to\infty, k\in K} (x^k, y^k) = (x^*, y^*)$ and $\lim_{k\to\infty, k\in K} (d^k, w^k) = (0, 0)$, where $\{(d^k, w^k)\}$ is the sequence generated in Step 3 of Algorithm 1.*

Proof (x^k, y^k) is updated to the $(\bar{x}^k, \bar{y}^k)$ from Step 1 and $(\bar{x}^k, \bar{y}^k)$ is updated to the next point (x^{k+1}, y^{k+1}) using Step 3. In case, if Step 1 does not update the iterative point then $(\bar{x}^k, \bar{y}^k) = (x^k, y^k)$ for that iteration. Since $\lim_{k\to\infty, k\in K} (x^k, y^k) = (x^*, y^*)$ therefore it holds that $\lim_{k\to\infty, k\in K} \{(\bar{x}^k, \bar{y}^k)\} = (x^*, y^*)$.

Since F is a directionally differentiable function, for $\epsilon > 0$ there exists $0 < \alpha_k \leq 1$ such that for each $\alpha \in (0, \alpha_k]$,

$$\left| \frac{F(\bar{x} + \alpha d^k, \bar{y} + \alpha w^k) - F(\bar{x}^k, \bar{y}^k)}{\alpha} - (\nabla_x F(\bar{x}^k, \bar{y}^k)^T d^k + \nabla_y F(\bar{x}^k, \bar{y}^k)^T w^k) \right| < \epsilon.$$

For $\epsilon = (\delta - 1)(\nabla_x F(\bar{x}^k, \bar{y}^k)^T d^k + \nabla_y F(\bar{x}^k, \bar{y}^k)^T w^k)$, and for each $\alpha \in (0, \alpha_k]$,

$$F(x^{k+1}, y^{k+1}) - F(\bar{x}^k, \bar{y}^k) \leq \delta\alpha_k(\nabla_x F(\bar{x}^k, \bar{y}^k)^T d^k + \nabla_y F(\bar{x}^k, \bar{y}^k)^T w^k). \quad (22)$$

The above inequality shows that the Armijo condition is satisfied for $\alpha_k > 0$. Since $F(\bar{x}^k, \bar{y}^k) < F(x^k, y^k)$, $F(x^{k+1}, y^{k+1}) - F(x^k, y^k) < F(x^{k+1}, y^{k+1}) - F(\bar{x}^k, \bar{y}^k)$. Hence (22) reduces to

$$F(x^{k+1}, y^{k+1}) - F(x^k, y^k) < \delta\alpha_k(\nabla_x F(\bar{x}^k, \bar{y}^k)^T d^k + \nabla_y F(\bar{x}^k, \bar{y}^k)^T w^k). \tag{23}$$

To prove the theorem, assume, on the contrary that a sub-sequence of $\{(d^k, w^k, \mu^k)\}, k \in K$ converges to a nonzero vector (d^*, w^*, μ^*). Without loss of generality, we can assume $(d^k, w^k, \mu^k) \to (d^*, w^*, \mu^*)$ and $\alpha_k \to \alpha_*$ for some subsequence K_1 of K. Then we claim that $\alpha_* > 0$. Again, on the contrary, assume that $\alpha_* = 0$. For large k, we have $J(\bar{\lambda}^k) \supseteq J(\lambda^*)$ and $I_{\bar{x}^k}(\bar{y}^k) \subseteq I_{x^*}(y^*)$. Hence it follows that if $W^k \to W^*$ then W^* satisfies $J(\lambda^*) \subseteq W^* \subseteq I_{x^*}(y^*)$. There exists a subsequnce K_2 of K_1 such that $k \to \infty, k \in K_2$ in the KKT optimality conditions of $QP_{(\bar{x}^k, \bar{y}^k)}(W^k)$, one may see that (d^*, w^*, μ^*) is the solution of $QP_{(x^*, y^*)}(W^*)$ for some W^* in $J(\lambda^*) \subseteq W^* \subseteq I_{x^*}(y^*)$. Here, $(0, 0, \lambda_{w^*})$ is feasible to $QP_{(x^*, y^*)}(W^*)$ with objective value 0. Furthermore, since objective function of $QP_{(x^*, y^*)}(W^*)$ is a strictly convex function and (d^*, w^*, μ^*) is optimal solution with $(d^*, w^*) \neq (0, 0)$, we have

$$(\nabla_x F(x^*, y^*)^T d^* + \nabla_y F(x^*, y^*)^T w^*) + \frac{1}{2}\begin{bmatrix} d^* \\ w^* \end{bmatrix}^T \begin{bmatrix} \nabla_{xx} F(x^*, y^*) & \nabla_{xy} F(x^*, y^*) \\ \nabla_{yx} F(x^*, y^*) & \nabla_{yy} F(x^*, y^*) \end{bmatrix} \begin{bmatrix} d^* \\ w^* \end{bmatrix} < 0.$$

This implies,

$$\nabla_x F(x^*, y^*)^T d^* + \nabla_y F(x^*, y^*)^T w^* < 0. \tag{24}$$

Since α_k is obtained using the backtracking process of the Armijo condition, stated in Algorithm 1, therefore for $\beta > \delta > 0$, $F(x^{k+1}, y^{k+1}) - F(\bar{x}^k, \bar{y}^k) > \delta\frac{\alpha_k}{\beta}(\nabla_x F(\bar{x}^k, \bar{y}^k) d^k + \nabla_y F(\bar{x}^k, \bar{y}^k) w^k)$. That is,

$$\frac{F(x^{k+1}, y^{k+1}) - F(\bar{x}^k, \bar{y}^k)}{\alpha_k} > \frac{\delta}{\beta}(\nabla_x F(\bar{x}^k, \bar{y}^k) d^k + \nabla_y F(\bar{x}^k, \bar{y}^k) w^k).$$

Taking limit $k \to \infty$, $(1 - \frac{\delta}{\beta})(\nabla_x F(x^*, y^*)^T d^* + \nabla_y F(x^*, y^*)^T w^*) \geq 0$, which contradicts (24) since $\delta < \beta$ in Algorithm 1. Therefore, $\alpha^* > 0$.

In limiting case, as $k \to \infty$, the right hand side of (23) is $\delta\alpha^*(\nabla_x F(x^*, y^*)^T d^* + \nabla_y F(x^*, y^*)^T w^*) < 0$, since $\alpha^* > 0$ and $\nabla_x F(x^*, y^*)^T d^* + \nabla_y F(x^*, y^*)^T w^* < 0$ from (24). This contradicts that the sequence $(x^{k+1}, y^{k+1}) = (\bar{x}^k, \bar{y}^k) + \alpha_k(d^k, w^k), k \in K$ converges to (x^*, y^*). Hence the result follows. ☐

Theorem 3.5 *Suppose the level set $L = \{(x, y) \mid F(x, y) \leq F(x^0, y^0)\}$ is bounded, where (x^0, y^0) is an initial point of Algorithm 1. Then,*

(i) If sequence generated by Algorithm 1 is infinite then each accumulation point of sequence $\{(x^k, y^k)\}$ satisfies the sufficient optimality condition of Theorem (3.1).
(ii) If Algorithm 1 terminates finitely then stopping point satisfies sufficient optimality condition of Theorem (3.1).

Proof If (x^k, y^k) is updated $(\bar{x}^k, \bar{y}^k)$ using Step 1 then $F(\bar{x}, \bar{y}) < F(x^k, y^k)$ at some kth iteration. Otherwise, if (x^k, y^k) is updated by Step 3, then using the arguments of Corollary (3.1), the strictly convex nature of F results $\nabla_x F(x^k, y^k)^T d^k + \nabla_y F(x^k, y^k)^T w^k < 0$. Therefore, F decreases in each iteration and the sequence $\{(x^k, y^k)\}$ lies in the level set L. From the boundness condition of the level set, there exists a subsequence $\{(x^k, y^k)\}, k \in K$ which converges to a unique accumulation point. Let (x^*, y^*) be the accumulation point of that subsequence.

Case (i): Since for large k, $J(\bar{\lambda}^k) \supseteq J(\lambda^*)$ and $I_{\bar{x}^k}(\bar{y}^k) \subseteq I_{x^*}(y^*)$ therefore W^* satisfies $J(\lambda^*) \subseteq W^* \subseteq I_{x^*}(y^*)$, where W^* is an accumulation point of the sequence $\{W^k\}$. The sequence $\{(d^k, w^k))\}, k \in K$ converges to $(0, 0)$ from Theorem 3.4. Using the KKT optimality condition on $QP_{(x^k, y^k)}(W^k)$ and taking limit $k \to \infty, k \in K'$ for some sub sequence of K, it can be concluded that $(0, 0, \lambda_{w^*})$ is the solution of $QP_{(x^*, y^*)}(W^*)$ for some λ_{w^*} corresponding to W^* satisfying $J(\lambda^*) \subseteq W^* \subseteq I_{x^*}(y^*)$. We claim that $(0, 0, \lambda_w)$ is solution of $QP_{(x^*, y^*)}(W)$ for some λ_w corresponding to each of W satisfying $J(\lambda^*) \subseteq W \subseteq I_{x^*}(y^*)$.

On the contrary, assume that there is some working set $\tilde{W}$ such that $\tilde{W} \neq W^*$ and $J(\lambda^*) \subseteq \tilde{W} \subseteq I_{x^*}(y^*)$ and $QP_{(x^*, y^*)}(\tilde{W})$ has nonzero descent direction vector $(\tilde{d}, \tilde{w})$. From Lemma 3.3, it follows that $(\bar{x}^k, \bar{y}^k)$ is projection of (x^k, y^k) on $S(I_{x^*}(y^*), J(\lambda^*))$ for some large k. Hence $I_{x^*}(y^*) \subseteq I_{\bar{x}^k}(\bar{y}^k)$ and $J(\bar{\lambda}^k) \subseteq J(\lambda^*)$ from the definition of $S(I_{x^*}(y^*), J(\lambda^*))$. Then, $J(\bar{\lambda}^k) \subseteq \tilde{W} \subseteq I_{\bar{x}^k}(\bar{y}^k)$ for some large

k. Let $(\tilde{d^k}, \tilde{w}^k, \tilde{\mu}^k)$ is the solution of $QP_{(\bar{x}^k,\bar{y}^k)}(\tilde{W})$. Without loss of generality, we can assume that $(\tilde{d^k}, \tilde{w}^k) \to (\tilde{d}, \tilde{w})$ as $k \to \infty$.

Now, for large $k \in K'$, the conditions $||\nabla_x F(x^k, y^k)d^k + \nabla_y F(x^k, y^k)w^k|| > \epsilon$ is satisfied as $\{(d^k, w^k)\} \to (0, 0)$ as $k \to \infty, k \in K'$. At this stage, (d^k, w^k) is the steepest descent direction due to Case (2b) in the method. However, for some large k, we can obtain

$$\nabla_x F(x^k, y^k)^T \tilde{d^k} + \nabla_y F(x^k, y^k)^T \tilde{w}^k < \nabla_x F(x^k, y^k)^T d^k + \nabla_y F(x^k, y^k)^T w^k, k \in \tilde{K},$$

due to the fact that $\{(d^k, w^k)\} \to (0, 0)$ and $\{(\tilde{d^k}, \tilde{w}^k)\} \to (\tilde{d}, \tilde{w}) \neq (0, 0)$. The above relation constitutes a contradiction, which validates our presumed claim.

Case (ii): Suppose the Algorithm 1 generates a finite sequence and stopped by Case (2c). Hence if z is the solution of the system $A^T_{W^*} z = -v(x^*, y^*)$ then the components associated with working set constraints $W^*/J(\lambda^*)$ and inequality constraints are positive. The active set matrix can be written as $A = \begin{bmatrix} A_{W^*} \\ A/A_{W^*} \end{bmatrix}$. Then,

$$[A^T_{W^*},\ (A/A_{W^*})^T] \begin{bmatrix} z \\ 0 \end{bmatrix} = -v(x^*, y^*).$$

Hence (18) can not hold at (x^*, y^*) for any appropriate $a, b, c \leq 0$, and for any solution of the system

$$A \begin{bmatrix} d' \\ w' \\ \mu' \end{bmatrix} = \begin{bmatrix} 0 \\ a \\ b \\ c \end{bmatrix}$$

corresponding to the active set matrix A. This means a nonzero descent direction does not exist at this stage by modifying the problem $QP_{(x^*,y^*)}(W)$ for any other W satisfying $J(\lambda^*) \subseteq W \subseteq I_{x^*}(y^*)$. As a result, the solution of $QP_{(x^*,y^*)}(W)$ is $(0, 0, \lambda_w)$ corresponding to every working set W satisfying $J(\lambda^*) \subseteq W \subseteq I_{x^*}(y^*)$. If (x^*, y^*) is determined by Case (2d), then the above argument holds true trivially. Thus for both of the cases, $(0, 0, \lambda_w)$ is solution of $QP_{(x^*,y^*)}(W)$ for each appropriate W.

Next we show that at (x^*, y^*) satisfies sufficient optimality condition of Theorem 3.1. Since F is a strict convex function, so $(d, w, \mu) = (0, 0, \lambda_w)$ is the global minimum point of $QP_{(x^*,y^*)}(W)$ for each $J(\lambda) \subseteq W \subseteq I_{x^*}(y^*)$, with 0 as optimal value. Since for any feasible point (d, w, μ) with $(d, w) \neq 0$, $\alpha(d, w, \mu)$ is also feasible, then for any $\alpha \in (0, 1]$, we have

$$\alpha \left[\nabla_x F(x^*, y^*)^T\ \nabla_y F(x^*, y^*)^T\right] \begin{bmatrix} d \\ w \end{bmatrix} + \frac{\alpha^2}{2} \begin{bmatrix} d \\ w \end{bmatrix}^T \begin{bmatrix} \nabla_{xx} F(x^*, y^*) & \nabla_{xy} F(x^*, y^*) \\ \nabla_{yx} F(x^*, y^*) & \nabla_{yy} F(x^*, y^*) \end{bmatrix} \begin{bmatrix} d \\ w \end{bmatrix} > 0. \quad (25)$$

Next, for $0 < \alpha \leq 1$,

$$F(x^* + \alpha d, y^* + \alpha w) = F(x^*, y^*) + \alpha \left[\nabla_x F(x^*, y^*)^T\ \nabla_y F(x^*, y^*)^T\right] \begin{bmatrix} d \\ w \end{bmatrix} + \frac{\alpha^2}{2} \begin{bmatrix} d \\ w \end{bmatrix}^T \begin{bmatrix} \nabla_{xx} F(x^*, y^*) & \nabla_{xy} F(x^*, y^*) \\ \nabla_{yx} F(x^*, y^*) & \nabla_{yy} F(x^*, y^*) \end{bmatrix} \begin{bmatrix} d \\ w \end{bmatrix} + o(\alpha^2).$$

Therefore there exists $\alpha_0 > 0$, so that $\forall \alpha \in (0, \alpha_0],\ (d, w, \mu) \in \Omega_{(x^*,y^*)}(W)$ with $\|d\| = 1$,

$$F(x^* + \alpha d, y^* + \alpha w) > F(x^*, y^*), \quad \text{(from (25))}$$

That is equivalent to $\nabla_x F(x^*, y^*)^T d + \nabla_y F(x^*, y^*)^T w \geq 0$ for all (d, w) in the set

$$\{(d, w, \mu) \mid (d, w, \mu) \in \bigcup_{J(\lambda^*) \subseteq W \subseteq I_{x^*}(y^*)} \Omega_{(x^*,y^*)}(W), \|d\| = 1\}.$$

Hence (x^*, y^*) satisfies the sufficient optimality condition of Theorem 3.1. □

Therefore, each accumulation point of the infinite sequence generated by the Algorithm 1 is an optimal solution to $CQBP$.

Remark 3.1 1. Algorithm 1 can obtain a stationary point if F is considered a convex function. In that case, if (x, y) is a local optimal solution of $CQBP$ then $F(x, y)^T d + \nabla_y F(x, y)^T w \geq 0$ for all $(d, w) \in \hat{C}(x, y)$. Hence if (d_W, w_W, μ_W) is the optimal solution of

$$\min_{d,w,\mu} \nabla_x F(x, y)^T d + \nabla_y F(x, y)^T w$$
$$\text{subject to } (d, w, \mu) \in \Omega_{(x,y)}(W), \ \|d\| \leq 1,$$

then $\min\{\nabla_x F(x, y)^T d_W + \nabla_y F(x, y)^T w_W \mid J(\lambda) \subseteq W \subseteq I_x(y)\} \geq 0$. This necessary condition is analogous to Theorem (5.4) of [5], which holds for the general case. The stopping point of Algorithm 1 satisfies the above necessary condition when F is a convex function.

2. The constraint qualification $LICQ$ is used on the lower level problem of $CQBP$, which may be relaxed by Mangasarian-Fromovitz Constraint Qualification ($MFCQ$) . In this case, the multipliers of the lower level problem of $CQBP$ may not be unique but lie in a bounded polyhedral set with the set of vertices $EU_x(y)$ say. Therefore contingent cone as in Section 5.3 of [5] becomes

$$\hat{C}(x, y) = \{(d, w) \mid (d, w, \mu) \in \bigcup_{\lambda \in EU_x(y)} \bigcup_{J(\lambda) \subseteq W \subseteq I_x(y)} \Omega_{(x,y)}(W)\}.$$

Hence the feasible direction can be obtained by solving a new subproblem $QP_{(x,y)}(W)$ for some W, $J(\lambda) \subseteq W \subseteq I_x(y)$, $\lambda \in EU_x(y)$. Accordingly, Algorithm 1 can be modified.

3. If the linear approximation of upper level objective function F is considered in place of quadratic approximation in the subproblem $QP_{(x,y)}(W)$, then the direction vector d can be restricted by $\|d\| \leq 1$ to avoid unboundedness.

4 Computational Experience

The steps of Algorithm 1 are explained in the following example and several numerical experiments are made later with different test problems. The numerical experiment section considers both quadratic and non-quadratic objective functions for the upper-level problem. We mention that even though Algorithm 1 has various nested loops, the method efficiently solves the $CQBP$ for some appropriately selected δ_0 and ϵ. In the next example, $CQBP$ with quadratic upper level is solved using the Algorithm 1 with detailed steps.

Example 4.1 *Consider the following $CQBP$*

$$\begin{array}{lll}
\min\limits_{x} & (x_1 - 4)^2 + (x_2 - 6)^2 + (y_1 - 4)^2 + (y_2 - 5)^2 & \\
\text{subject to} & -x_1 - 2y_2 + 28 \leq 0 & \\
& -y_1 \leq -8 & \\
& -x_1 - x_2 \leq -20 & \\
& \min\limits_{y} \quad (x_1 - y_1)^2 + (x_2 - y_2)^2 & \\
& \text{subject to} \quad -y_1 \leq 0, & (I) \\
& \qquad -y_2 \leq 0, & (II) \\
& \qquad y_1 \leq 10, & (III) \\
& \qquad y_2 \leq 10. & (IV)
\end{array}$$

Here, $F(x, y) = (x_1 - 4)^2 + (x_2 - 6)^2 + (y_1 - 4)^2 + (y_2 - 5)^2$, which is a convex quadratic function. Hence by Theorem 3.3 step length $\alpha_k = 1$ for every iteration k. Let the initial point be $x^0 = \begin{bmatrix} 11 \\ 12 \end{bmatrix}$. At x^0, the optimal solution of the lower level problem is computed as $y^0 = y(x^0) = \begin{bmatrix} 10 \\ 10 \end{bmatrix}$. Let the initial values of parameter $\delta_0 = 0.1$ and $\epsilon = 10^{-4}$.

Iteration 0: The corresponding working set at (x^0, y^0) is $W^0 = I_{x^0}(y^0) = \{III, IV\}$.

Here, $I_0(\delta') = I_{x^0}(y^0)$ and $J(\lambda^0) = J_0(\delta') = \emptyset$. In Step 2, at this iteration $QP_{(x^0,y^0)}(W^0)$ can be formulated as

$$QP_{(x^0,y^0)}(W^0): \quad \begin{array}{ll} \min\limits_{d,w,\mu} & d_1^2 + d_2^2 + w_1^2 + w_2^2 + 14d_1 + 12d_2 + 12w_1 + 10w_2 \\ \text{subject to} & -d_1 - 2w_2 \le 3, \\ & -w_1 \le 2, -d_1 - d_2 \le 3, \\ & -2d_1 + 2w_1 + \mu_3 = 2, \\ & -2d_2 + 2w_2 + \mu_4 = 4, \\ & w_1 = 0, w_2 = 0, -\mu_3 \le 0, -\mu_4 \le 0. \end{array}$$

The solution of this subproblem is $(d^0, w^0, \mu^0) = \left(\begin{bmatrix} -1 \\ -2 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right)$.
So $||\nabla_x F(x^0, y^0)d^0 + \nabla_y F(x^0, y^0)w^0|| > \epsilon$.

Iteration 1: Next iterative point is $(x^1, y^1) = (x^0, y^0) + (d^0, w^0) = \left(\begin{bmatrix} 10 \\ 10 \end{bmatrix}, \begin{bmatrix} 10 \\ 10 \end{bmatrix} \right)$. From Theorem (3.2), $y^1 = y(x^1)$. One can verify that y^1 is the optimal solution of the lower level problem at x^1. Here, $T^0 = \phi$. Hence the working set is updated as $W^1 = W^0 \cup T^0 = \{III, IV\}$. At (x^1, y^1), we have $I_{x^1}(y^1) = I_1(\delta'),\ J(\lambda^1) = J_1(\delta')$.

Iteration-2: The subproblem $QP_{(x^1,y^1)}(W^1)$ is

$$QP_{(x^1,y^1)}(W^1): \quad \begin{array}{ll} \min\limits_{d,w,\mu} & d_1^2 + d_2^2 + w_1^2 + w_2^2 + 12d_1 + 8d_2 + 12w_1 + 10w_2 \\ \text{subject to} & -d_1 - 2w_2 \le 2, \\ & -w_1 \le 2, -d_1 - d_2 \le 0, \\ & -2d_1 + 2w_1 + \mu_3 = 0, \\ & -2d_2 + 2w_2 + \mu_4 = 0, \\ & w_1 = 0, w_2 = 0, -\mu_3 \le 0, -\mu_4 \le 0. \end{array}$$

The solution of the subproblem is $(d, w, \mu) = \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right)$. Since $||\nabla_x F(x^1, y^1)d + \nabla_y F(x^1, y^1)w|| \le \epsilon$, so either (x^1, y^1) is the solution or the working set has to be modified. Here

$$A_{W^1} = \begin{bmatrix} -2 & 0 & 2 & 0 & 1 & 0 \\ 0 & -2 & 0 & 2 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \end{bmatrix}, v(x^1, y^1) = -\begin{bmatrix} 12 \\ 8 \\ 12 \\ 10 \\ 0 \\ 0 \end{bmatrix}.$$

The solution of $A_{W^1}^T z = -v(x^1, y^1)$ is $z^T = (2, 0, -16, -10, 8, 2)$. Here, all the components of z corresponding to the working set are not positive. Hence (x^1, y^1) is not the stopping point, and the working set has to be modified.

$$\{z_s \mid \lambda_{i_s} = 0, , i_s \in W^2\} = \{z_3, z_4\} = \{-16, -10\}.$$

The minimum of this set occurs corresponding to the constraint in the working set
$i_3 = \{III\}$. Hence the modified working set is $W^1 \longleftarrow W^1/\{III\} = \{IV\}$. Accordingly $QP_{(x^1,y^1)}(W^1)$ is modified with this new working set W^1 to get a nonzero step.

$$QP_{(x^1,y^1)}(W^1): \quad \begin{array}{ll} \min\limits_{d,w,\mu} & d_1^2 + d_2^2 + w_1^2 + w_2^2 + 12d_1 + 8d_2 + 12w_1 + 10w_2 \\ \text{subject to} & -d_1 - 2w_2 \le 2, w_1 \le 0, \\ & -w_1 \le 2, -d_1 - d_2 \le 0, \\ & -2d_1 + 2w_1 = 0, \\ & -2d_2 + 2w_2 + \mu_4 = 0, \\ & w_2 = 0, -\mu_4 \le 0. \end{array}$$

The solution of this subproblem is $(d^1, w^1, \mu^1) = \left(\begin{bmatrix} -2 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \end{bmatrix}, 4 \right)$.
Here, we have $||\nabla_x F(x^1, y^1)d^1 + \nabla_y F(x^1, y^1)w^1|| > \epsilon$. Hence the next point can be computed as

$(x^2,y^2)=(x^1,y^1)+(d^1,w^1)=\left(\begin{bmatrix}8\\12\end{bmatrix},\begin{bmatrix}8\\10\end{bmatrix}\right)$. Since the set $T^1=\phi$, therefore $W^2=W^1\cup T^1=\{IV\}$. At (x^2,y^2), we have $I_{x^2}(y^2)=I_2(\delta')$, $J(\lambda^2)=J_2(\delta')$. The next iterating point (x^3,y^3) is to be determined.
Iteration-3: The subproblem corresponding to the working set W^2 is

$$QP_{(x^2,y^2)}(W^2):\quad \begin{aligned} \min_{d,w,\mu}\quad & d_1^2+d_2^2+w_1^2+w_2^2+8d_1+12d_2+8w_1+10w_2\\ \text{subject to}\quad & -d_1-2w_2\leq 0,\\ & -w_1\leq 2, -d_1-d_2\leq 0,\\ & -2d_1+2w_1=0,\\ & -2d_2+2w_2+\mu_4=4,\\ & w_2=0, -\mu_4\leq 0.\end{aligned}$$

The solution of this subproblem is $(d,w,\mu)=\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}0\\0\end{bmatrix},4\right)$. Here $(d,w)=(0,0)$. Hence as in iteration 2, the matrices A_{W^2} are computed.

$$A_{W^2}=\begin{bmatrix}-2&0&2&0&0\\0&-2&0&2&1\\0&0&0&1&0\\-1&0&0&-2&0\\0&0&0&0&-1\end{bmatrix},\quad v(x^2,y^2)=-\begin{bmatrix}8\\12\\8\\10\end{bmatrix}.$$

Solution of $A_{W^2}^Tz=-v(x^2,y^2)$ is $z^T=(-4,6,10,16,6)$. Since $z_s\geq 0$ corresponds to the inequality constraints and working set constraints, therefore (x^2,y^2) is a stopping point of Algorithm 1. Hence $(x^*,y^*)=(x^2,y^2)=([8,12]^T,[8,10]^T)$ is the local optimal solution.
Note that (x^*,y^*) satisfies the stopping criteria mentioned in Case (2c) corresponding to the working set $W^2=\{IV\}$. In this final iteration, $W^*=W^2=\{IV\}$ is the optimal working subset, $(d^2,w^2)=(0,0)$, $\lambda^*=\lambda^2=(0,0,0,4)$, and $\lambda_{w^*}=\mu^2=4$. □

In the following example, we describe the use of delta active search technique.

Example 4.2

$$\begin{aligned} \min_{x,y}\quad & -(x-1.2)^5-(y-1.2)^5\\ \textit{subject to}\quad & 0\leq x\leq 1.2,\\ & \quad \textit{min}\quad y^2-2xy\\ & \quad \textit{subject to} -y\leq 0, \qquad (I)\\ & \qquad\qquad -x-y+2\leq 0. \qquad (II)\end{aligned}$$

Consider the value $\delta_0=0.1$ and $\epsilon=10^{-4}$. The initial point $(x^0,y^0)=(0.8,1.2)$ is updated to (x^1,y^1), and then (x^1,y^1) is updated to (x^2,y^2) using Step 3 of Algorithm 1 as follows

$$\begin{aligned}(x^1,y^1)&=(x^0,y^0)+(d^0,w^0)=(0.8,1.2)+(0.1,-0.1)=(0.9,1.1),\\ (x^2,y^2)&=(x^1,y^1)+(d^1,w^1)=(0.9,1.1)+(0.0714,-0.0714)=(0.9714,1.0286).\end{aligned}$$

At this point, we have $\lambda^2=(0,0.1144)^T$. Here,
$I_{x^2}(y^2)=\{II\}$, $J(\lambda^2)=\{II\}$, $J_2(\delta')=\emptyset$ and $I_2(\delta')=\{II\}$. Since $J_2(\delta')\neq(J(\lambda^2)$, from Step 1, we project (x^2,y^2) on the set $S(I_2(\delta'),J_2(\delta'))$. For this, the following problem is solved

$$\begin{aligned} \min\quad & d^2+w^2\\ \text{subject to}\quad & 0\leq x^2+d\leq 1.2\\ & 2(y^2+w)-2(x^2+d)-(\lambda_2^2+\mu_2)=0,\\ & -(x^2+d)-(y^2+w)+2=0,\\ & -y^2-w\leq 0,\\ & \lambda_2^2+\mu_2=0.\end{aligned}$$

We obtain the solution $(\bar{d}, \bar{w}, \bar{\mu}_2) = (0.0286, -0.0286, -0.1144)$. Hence the projection of (x^2, y^2) on the set $S(I_2(\delta'), J_2(\delta'))$ is computed as $(\bar{x}, \bar{y}) = (x^2, y^2) + (\bar{d}, \bar{w}) = (1, 1)$ and $\bar{\lambda} = (0, 0)^T$. We accept this projection as $F(\bar{x}, \bar{y}) < F(x^2, y^2)$. At the new point $(1, 1)$, again from Step 3, we update the working set as $W' = W^0 - \{II\} = \emptyset$. With this modification, the sequence of iterative points (x^k, y^k) converges to $(1.2, 1.2)$ using Step 3 of Algorithm 1. Thus $(1.2, 1.2)$ is the solution of the given problem. □

In the next section, we verify the methodology for a set of test problems.

4.1 Numerical Experiment with test problems

Here, we provide the computational details of the implementation of Algorithm 1. MATLAB code is developed for Algorithm 1 in MATLAB-2023 and implemented on some test problems, which are collected from [29], whose upper level objective functions are both quadratic and nonquadratic type and lower levels are quadratic problems.

It is not possible to obtain an exact zero in the code. Therefore, the following accuracy is accepted. Active constraint tolerance is 10^{-4}. Optimality tolerance is $\|(\nabla_x F(x^*, y^*)d + \nabla_y F(x^*, y^*)w\| \leq 10^{-6}$. We have considered the following values for the various initial parameters in Algorithm 1:

$$\delta = 10^{-3},\ \delta_0 = 0.1, \epsilon = 10^{-4}, \sigma_1 = \sigma_2 = 0.5.$$

The subproblems are solved using QUADPROG in MATLAB-2023. The computational results are summarized in the table 4.1. One may observe in the table that the solution obtained by the proposed Algorithm 1 is the same as that provided in ([29]), which justifies the validity of this methodology.

Table 1 Numerical experiment with test problems

Sr. N.	Problem Description	Initial Point	Optimal solution
1	AiyoshiShimizu1984Ex2	[0 10 -10 -10]	[0 10^{-15} -10 -10]
2	Bard1988Ex1	[2 2.5]	[1 0]
3	Bard1991Ex1	[2 0 6]	[2 6 0]
4	DeSilva1978	[0 0 0.5 0.5]	[0.5 0.5 0.5 0.5]
5	FalkLiu1995	[0 0 0.5 0.5]	[0.75 0.75 0.75 0.75]
6	FloudasEtal2013	[0 10 -10 -10]	[0 0 -10 -10]
7	GumusFloudas2001Ex4	[2 5]	[3 5]
8	HatzEtal2013	[2 2 0]	[0 0 0]
9	HendersonQuandt1958	[0 50]	[93.3333 26.6667]
10	LamparielloSagratella2017Ex31	[2 -1]	[1 0]
11	MuuQuy2003Ex1	[0 0 0]	[0.8462,0.7692,0]
12	MuuQuy2003Ex2	[0 0 0 0 2]	[0.6111 0.3889 0.0000 0.0000 1.8333]
13	Outrata1990Ex1a	[0 0 0 0]	[1.0316 3.0978 2.5970 1.7929]
14	Outrata1990Ex1b	[0 0 0 0]	[0.2788 0.4748 2.3438 1.0325]
15	Outrata1990Ex1c	[0 0 0 0]	[11.9940 38.9805 2.9985 2.9985]
16	Outrata1990Ex1d	[1 1 1 0]	[2 0 2 0]
17	Outrata1990Ex1e,	[1 1 1 0]	[-0.4000 0.8000 2.0000 0.0000]
18	Outrata1990Ex2a	[0 2.0998 0.2998]	[2.6624 2.9985 2.9985]
19	ShimizuAiyoshi1981Ex1	[15 5]	[10 10]
20	ShimizuAiyoshi1981Ex2	[0 15 0 10]	[20 5 10 5]
21	ShimizuEtal1997a	[1 0]	[0.9459 -0.1622]
22	ShimizuEtal1997b^a	[4 16]	[7.2000 12.8000]
23	SinhaMaloDeb2014TP6	[1.5 1.1951 0.0061]	[1.8889 0.888 0.0000]
24	TuyEtal2007	[0,5]	[1.5 4.5]
25	Yezza1996Ex41	[0,0]	[3 1]

5 Conclusion

This paper uses an active stage strategy for a particular class of nonlinear bilevel programming problems ($CQBP$) to generate a descent sequence converging to the local optimal point. Quadratic approximation of the upper level objective ensures fast convergence to the optimal solution. The methodology can address coupling constraints at the upper level problem of $CQBP$. This iterative process can be further developed for general nonlinear bilevel programming problems by modifying the subproblem at every iteration, which is the future scope of the present contribution.

Acknowledgments The work of first author is supported by Council of Scientific & Industrial Research, India (Grant no: 09/081(1400)/2019-EMR-I).

Declarations

Conflict of interest The authors declare that they have no conflicts of interest to this work.

Data availability Data sharing was not applicable to this article as no data sets were generated or analysed during the current study.

References

1. Bard, J.: Practical Bilevel Optimization: Algorithms and Applications. Nonconvex Optimization and Its Applications. Springer US (1998). URL https://books.google.co.in/books?id=3T9LZreZshUC
2. Bard, J.F., Moore, J.T.: A branch and bound algorithm for the bilevel programming problem. SIAM Journal on Scientific and Statistical Computing **11**(2), 281–292 (1990). URL https://doi.org/10.1137/0911017
3. Calvete, H.I., Gale, C., Mateo, P.M.: A new approach for solving linear bilevel problems using genetic algorithms. European Journal of Operational Research **188**(1), 14–28 (2008). URL https://doi.org/10.1016/j.ejor.2007.03.034
4. Dempe, S.: A necessary and a sufficient optimality condition for bilevel programming problems. Optimization **25**(4), 341–354 (1992). URL https://doi.org/10.1080/02331939208843831
5. Dempe, S.: Foundations of bilevel programming. dordrecht, the netherlands: Kluwer academic publishers (2002)
6. Dempe, S., Dutta, J.: Is bilevel programming a special case of a mathematical program with complementarity constraints? Mathematical programming **131**(1), 37–48 (2012). URL https://doi.org/10.1007/s10107-010-0342-1
7. Dempe, S., Dutta, J., Mordukhovich, B.: New necessary optimality conditions in optimistic bilevel programming. Optimization **56**(5-6), 577–604 (2007). URL https://doi.org/10.1080/02331930701617551
8. Dempe, S., Zemkoho, A.B.: The bilevel programming problem: reformulations, constraint qualifications and optimality conditions. Mathematical Programming **138**(1), 447–473 (2013). URL https://doi.org/10.1007/s10107-011-0508-5
9. Etoa, J.B.E.: Solving convex quadratic bilevel programming problems using an enumeration sequential quadratic programming algorithm. Journal of Global Optimization **47**(4), 615–637 (2010). URL https://doi.org/10.1007/s10898-009-9482-3
10. Facchinei, F., Jiang, H., Qi, L.: A smoothing method for mathematical programs with equilibrium constraints. Mathematical programming **85**(1), 107 (1999). URL https://doi.org/10.1007/s10107990015a
11. Gill, P.E., Wong, E.: Methods for convex and general quadratic programming. Mathematical programming computation **7**(1), 71–112 (2015). URL https://doi.org/10.1007/s12532-014-0075-x
12. Han, J., Zhang, G., Hu, Y., Lu, J.: A solution to bi/tri-level programming problems using particle swarm optimization. Information Sciences **370**, 519–537 (2016). URL https://doi.org/10.1016/j.ins.2016.08.022
13. Jiang, H., Ralph, D.: Smooth sqp methods for mathematical programs with nonlinear complementarity constraints. SIAM Journal on Optimization **10**(3), 779–808 (2000). URL https://doi.org/10.1137/S105262349733232
14. Kanzow, C., Schwartz, A.: A new regularization method for mathematical programs with complementarity constraints with strong convergence properties. SIAM Journal on Optimization **23**(2), 770–798 (2013). URL https://doi.org/10.1137/100802487
15. Kuo, R., Huang, C.: Application of particle swarm optimization algorithm for solving bi-level linear programming problem. Computers & Mathematics with Applications **58**(4), 678–685 (2009). URL https://doi.org/10.1016/j.camwa.2009.02.028
16. Luo, Z.Q., Pang, J.S., Ralph, D.: Mathematical programs with equilibrium constraints. Cambridge University Press (1996)
17. Mersha, A.G., Dempe, S.: Feasible direction method for bilevel programming problem. Optimization **61**(5), 597–616 (2012). URL https://doi.org/10.1080/02331934.2011.641546
18. Nocedal, J., Wright, S.J.: Numerical optimization. Springer New York, NY (1999). URL https://doi.org/10.1007/978-0-387-40065-5
19. Savard, G., Gauvin, J.: The steepest descent direction for the nonlinear bilevel programming problem. Operations Research Letters - ORL **15**, 265–272 (1994). DOI 10.1016/0167-6377(94)90086-8
20. Sinha, A., Malo, P., Deb, K.: Evolutionary algorithm for bilevel optimization using approximations of the lower level optimal solution mapping. European Journal of Operational Research **257**(2), 395–411 (2017). URL https://doi.org/10.1016/j.ejor.2016.08.027
21. Sinha, A., Malo, P., Deb, K.: A review on bilevel optimization: from classical to evolutionary approaches and applications. IEEE Transactions on Evolutionary Computation **22**(2), 276–295 (2017). URL https://10.1109/TEVC.2017.2712906
22. Sinha, A., Soun, T., Deb, K.: Using karush-kuhn-tucker proximity measure for solving bilevel optimization problems. Swarm and evolutionary computation **44**, 496–510 (2019). URL https://doi.org/10.1016/j.swevo.2018.06.004
23. Steffensen, S., Ulbrich, M.: A new relaxation scheme for mathematical programs with equilibrium constraints. SIAM Journal on Optimization **20**(5), 2504–2539 (2010). URL https://doi.org/10.1137/090748883

24. Wang, Y., Jiao, Y.C., Li, H.: An evolutionary algorithm for solving nonlinear bilevel programming based on a new constraint-handling scheme. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) **35**(2), 221–232 (2005). URL https://doi.org/10.1109/TSMCC.2004.841908
25. Wang, Y., Li, H., Dang, C.: A new evolutionary algorithm for a class of nonlinear bilevel programming problems and its global convergence. INFORMS Journal on Computing **23**(4), 618–629 (2011). URL https://doi.org/10.1287/ijoc.1100.0430
26. Ye, J., Zhu, D.: Optimality conditions for bilevel programming problems. Optimization **33**(1), 9–27 (1995). URL https://doi.org/10.1080/02331939508844060
27. Ye, J.J., Zhu, D.: New necessary optimality conditions for bilevel programs by combining the mpec and value function approaches. SIAM Journal on Optimization **20**(4), 1885–1905 (2010). URL https://doi.org/10.1137/080725088
28. Zhang, J.Z., Liu, G., Wang, S.: A globally convergent approximately active search algorithm for solving mathematical programs with linear complementarity constraints. Numerische Mathematik **98**(3), 539–558 (2004). URL https://doi.org/10.1007/s00211-004-0542-9
29. Zhou, S., Zemkoho, A.B., Tin, A.: Bolib: Bilevel optimization library of test problems. In: Bilevel Optimization, pp. 563–580. Springer (2020). URL https://doi.org/10.1007/978-3-030-52119-6_19