A SUBGRADIENT-BASED CONVEX APPROXIMATIONS METHOD FOR DC PROGRAMMING AND ITS APPLICATIONS

JIAN GU
School of Sciences
Dalian Ocean University
Dalian 116023, China

XIANTAO XIAO AND LIWEI ZHANG
School of Mathematical Sciences
Dalian University of Technology
Dalian 116023, China

(Communicated by Nguyen Van Thoai)

Abstract. We consider an optimization problem that minimizes a function of the form
\[ f = f_0 + f_1 - f_2 \]
with the constraint \( g - h \leq 0 \), where \( f_0 \) is continuous differentiable, \( f_1, f_2 \) are convex and \( g, h \) are lower semicontinuous convex. We propose to solve the problem by an inexact subgradient-based convex approximations method. Under mild assumptions, we show that the method is guaranteed to converge to a stationary point. Finally, some preliminary numerical results are given.

1. Introduction. Let \( \mathcal{X} \) be a finite dimensional real Hilbert space equipped with an inner product \( \langle \cdot, \cdot \rangle \) and its induced norm \( \| \cdot \| \). In this paper, we consider the following nonlinear optimization problem:
\[
\begin{align*}
\min & \quad f(x) = f_0(x) + f_1(x) - f_2(x) \\
\text{s.t.} & \quad g(x) - h(x) \leq 0, \\
& \quad x \in K,
\end{align*}
\]
where \( f_0 : \mathcal{X} \to \mathbb{R} \) is continuous differentiable, \( f_i : \mathcal{X} \to \mathbb{R}, i = 1, 2 \) are convex, \( g = (g_1, g_2, \ldots, g_p)^T : \mathcal{X} \to \mathbb{R}^p \) and \( h = (h_1, h_2, \ldots, h_p)^T : \mathcal{X} \to \mathbb{R}^p \) are lower semicontinuous convex, and \( K \subseteq \mathcal{X} \) is a closed convex set. The functions \( f_1(x) - f_2(x) \) and \( g(x) - h(x) \) are usually called DC functions (difference of two convex functions), and the problem (1) is called DC programming (DCP). It is clear that if \( f_0 \equiv 0 \) and both \( f_2(x) \) and \( h(x) \) are affine functions, problem (1) is a convex optimization problem. In applications, there are many problems that either can be directly transformed into the form of DCP or can be solved by a sequence of DCP approximations. We list some of them in the following. For an introduction to the theory and applications of DCP, readers are referred to the survey \cite{11} and the references \cite{6, 22, 2, 3}.

2010 Mathematics Subject Classification. Primary: 90C26, 49M05; Secondary: 49J53.

Key words and phrases. DC programming, convex program, stationary point, set-valued analysis, convergence.

The work was supported by the National Natural Science Foundation of China (11071029 and 11101064) and Doctoral Fund of Ministry of Education of China (20110041120039).
1.1. Quandratically constrained quadratic programming. Given a group of quadratic functions:

\[ q_i(x) := \frac{1}{2} \langle A_i x, x \rangle + \langle b^i, x \rangle + r_i, \quad i = 0, 1, \ldots, m, \]

where \( A_i \) is an \( n \times n \) symmetric matrix, \( b^i \in \mathbb{R}^n \) and \( r_i \in \mathbb{R}, \quad i = 0, 1, \ldots, m \). The problem named as quadratically constrained quadratic programming (QCQP) is in the following form:

\[
\begin{aligned}
\min & \quad q_0(x) \\
\text{s.t.} & \quad q_i(x) \leq 0, \quad i = 1, \ldots, m, \\
& \quad x \in \mathbb{R}^n.
\end{aligned}
\] (2)

If some of \( A_i, i = 0, 1, \ldots, m \) are not positive semidefinite, problem (2) is nonconvex and difficult to be solved. The nonconvex problem has been intensively studied, such as in [1, 4, 21, 13, 9, 12].

In [1] the author pointed out that the quadratic function can be equivalently written as a DC function. For \( i \in \{0, 1, \ldots, m\} \), let \( \lambda^i \) be the largest eigenvalue of \( A_i \) and \( \rho^i \) be a number satisfying \( \rho^i > \max\{0, \lambda^i\} \). Define

\[ g_i(x) := \frac{1}{2} \rho^i \|x\|^2 + r_i \]

and

\[ h_i(x) := \frac{1}{2} \langle (\rho^i I - A_i) x, x \rangle - \langle b^i, x \rangle, \]

where \( I \) denotes the identity matrix. Note that both \( g_i \) and \( h_i \) are convex and \( \lambda^i \) is a DC function. Thus, the problem (2) can be written in the form of DCP.

1.2. Low rank matrix optimization. Consider the following matrix optimization problem:

\[
\begin{aligned}
\min & \quad f(X) \\
\text{s.t.} & \quad g(X) \leq 0, \\
& \quad \text{rank}(X) \leq r, \\
& \quad X \in Q,
\end{aligned}
\] (3)

where \( f \) is continuous differentiable, \( g \) is convex and the set \( Q \subseteq \mathbb{R}^{m \times n} \) is closed convex. Without loss of generality, we assume \( m \leq n \). The constraint rank(\( X \)) \( \leq r \) is called low rank constraint since \( r > 0 \) is very small in applications. Recently, the low rank matrix optimization problem is a hot research topic in many fields, such as applied statistics, image processing, risk management (see [7, 16, 8]).

The low rank constraint is nonconvex and can be written in a DC form. Given a matrix \( X \in \mathbb{R}^{m \times n} \), let \( \sigma_1(X) \geq \sigma_2(X) \geq \cdots \geq \sigma_m(X) \geq 0 \) be the singular values of \( X \) arranged in the non-increasing order, the Ky Fan \( k \)-norm \( \|X\|_{(k)} \) of \( X \) is defined as the sum of its \( k \) largest singular values, i.e.,

\[ \|X\|_{(k)} := \sum_{i=1}^{k} \sigma_i(X), \]

where \( 1 \leq k \leq m \). Note that the constraint rank(\( X \)) \( \leq r \) is satisfied if and only if

\[ \sigma_{r+1}(X) + \sigma_{r+2}(X) + \cdots + \sigma_m(X) = \|X\|_{(m)} - \|X\|_{(r)} = 0. \]
Therefore, it is possible to solve problem (3) by considering the penalized problem:

$$\min \quad f(X) + c\|X\|_{(m)} - c\|X\|_{(r)}$$
\[
\text{s.t.} \quad g(X) \leq 0, \tag{4}
X \in Q.
\]

Since $\|X\|_{(k)}$ is convex in $X$, the above problem is also a type of DCP.

1.3. Chance constrained programs. In stochastic programming community, solving chance constrained programs continue to be an attractive topic since 1960s. For a survey of the topic, readers are referred to Chapter 4 of [20] and references therein. The problem is usually formulated as:

$$\min \quad f(x)$$

$$\text{s.t.} \quad \Pr\{c(x, \xi) \leq 0\} \geq 1 - \alpha, \tag{5}
x \in Q,$$

where $\xi$ is an $m$-dimensional random vector, $Q \subseteq \mathbb{R}^n$ is a closed convex set, $f : \mathbb{R}^n \to \mathbb{R}$ is a real-valued function and $c : \mathbb{R}^{n+m} \to \mathbb{R}$ is convex in $x$. In problem (5), the constraint $c(x, \xi) \leq 0$ is required to be satisfied with a probability at least $1 - \alpha$. The constraint

$$\Pr\{c(x, \xi) \leq 0\} \geq 1 - \alpha \tag{6}$$

is called chance constraint (also called probabilistic constraint). Owing to the non-convex characteristic, chance constrained programs are computationally intractable and have motivated the need for approximation solution schemes.

Let $1_A(\cdot)$ be the indicator function of set $A$ as:

$$1_A(z) = \begin{cases} 
1, & \text{if } z \in A, \\
0, & \text{otherwise.}
\end{cases}$$

Let $Z = c(x, \xi)$, the chance constraint (6) can be rewritten in the form of:

$$\Pr\{Z > 0\} \leq \alpha,$$

which is equivalent to an expectation constraint,

$$\mathbb{E}[1_{(0, +\infty)}(Z)] \leq \alpha.$$

Since $1_{(0, +\infty)}(z)$ is clearly nonconvex, one way to approximate the chance constraint is to find a convex approximation $g(z)$ of $1_{(0, +\infty)}(z)$ such that $g(z) \geq 1_{(0, +\infty)}(z)$ for any $z \in \mathbb{R}$. Then, $\mathbb{E}[g(Z)] \leq \alpha$ would be a convex conservative approximation of constraint (6). In [15], the CVaR approximation is proved to be the best conservative convex approximation, which uses

$$g(z) = \frac{1}{t}[t + z]_+$$
to approximate $1_{(0, +\infty)}(z)$, where $t > 0$ is a parameter and $[z]_+ = \max\{0, z\}$. And the approximated constraint is formed as

$$\inf_{t>0} \{t^{-1}\mathbb{E}[t + Z]_+ - \alpha\} \leq 0,$$

which is equivalent to

$$\inf_{t \in \mathbb{R}} \{\alpha^{-1}\mathbb{E}[Z - t]_+ + t\} \leq 0.$$

The left hand side of the above inequality is well known as conditional value at risk (CVaR), i.e.,

$$\text{CVaR}_{1-\alpha}(Z) := \inf_{t \in \mathbb{R}} \{\alpha^{-1}\mathbb{E}[Z - t]_+ + t\}.$$
Figure 1. $1_{(0, +\infty)}(z)$, CVaR approximation and DC approximation

From Figure 1 we can see that, although CVaR approximation is the best convex conservative approximation, it is not a good approximation. In [10, 23, 19], the DC approximation is applied by using

$$g(z) = \frac{1}{t}[t + z]_+ - \frac{1}{t}[z]_+$$

to approximate $1_{(0, +\infty)}(z)$. When $t > 0$ is small enough, it is clearly a much better approximation. Then the chance constraint (6) is approximated by

$$\frac{1}{t}E[t + Z]_+ - \frac{1}{t}E[Z]_+ - \alpha \leq 0,$$

and the corresponding approximation problem of (5) is in the same formulation as problem (1) (when $t$ is fixed).

All these examples demonstrate that the demand for efficient approaches to solve DCP grows rapidly. Due to the nonconvex nature (maybe also nonsmooth), there are few efficient approaches proposed in the literature to address DCP. In [11, 22], a combinational method of branch and bound type was established for globally solving the DC programs. One obvious discouraging fact for the branch and bound type method is that it usually needs a lot of iterations to obtain a solution, which makes it powerless for large scale problems. A local method called DCA was studied in [3], which solves the problem for minimizing a DC function on the whole space. Given a good starting point, DCA was proved to converge to a global solution and be very efficient even for the large scale problems. In [1], the author also gave a combined DCA-branch-and-bound algorithm for minimizing a quadratic function under convex quadratic constraints which is in the form of (2).

In this paper, we consider a more general formulation in which the DC term appears both in the object function and in the constraint. We propose an inexact subgradient-based convex approximations method for solving problem (1). The basic idea is that, by linearizing the second term of the DC function, DCP can be approximated by a convex program. We present an iteration method in which a sequence of convex subproblems is solved. We show that the sequence of solutions converge to a stationary point of the DCP. We also report some preliminary numerical results of the method for solving joint chance constrained programs and quadratically constrained quadratic programs. In theory the method only guarantees to get a stationary point, but the numerical results show that for most case we can obtain the global solution as long as a suitable initial point is offered.

The rest of the paper is organized as follows. We give some preliminaries in Section 2. In Section 3, the inexact subgradient-based convex approximations method is presented and its convergence results are established. The numerical results are reported in Section 4.
2. Preliminaries. For the sake of subsequent discussions, we introduce some related concepts in this section. Let $\mathcal{X}$ and $\mathcal{Y}$ be finite dimensional real Hilbert spaces.

**Definition 2.1.** [5, Definition 2.44] A mapping $g : \mathcal{X} \to \mathcal{Y}$ is said to be directionally differentiable at a point $x \in \mathcal{X}$ in a direction $d \in \mathcal{X}$, if the limit

$$g'(x; d) := \lim_{t \downarrow 0} \frac{g(x + td) - g(x)}{t}$$

exists. If $g$ is directionally differentiable at $x$ in every direction $d \in \mathcal{X}$, we say that $g$ is directionally differentiable at $x$.

**Definition 2.2.** [14] Let $f : \mathcal{X} \to (-\infty, +\infty]$ be a Lipschitz continuous function.

(i) Clarke’s generalized directional derivative of $f$ at $\bar{x} \in \mathcal{X}$ in direction $d \in \mathcal{X}$ is defined by

$$f^\circ(\bar{x}; d) := \limsup_{t \downarrow 0} \frac{f(x + td) - f(x)}{t}.$$  

(ii) Clarke’s subdifferential (or generalized gradient) of $f$ at $\bar{x}$ is defined by

$$\partial f(\bar{x}) := \{v \in \mathcal{X}^* | \langle v, d \rangle \leq f^\circ(\bar{x}; d) \text{ for all } d \in \mathcal{X}\},$$

where $\mathcal{X}^*$ is the dual space of $\mathcal{X}$.

(iii) We say $f$ is Clarke regular at $\bar{x}$, if $f$ is directionally differentiable at $\bar{x}$ and $f'(\bar{x}; d) = f^\circ(\bar{x}; d)$ for all $d \in \mathcal{X}$.

**Definition 2.3.** Let $S \subset \mathcal{X}$ be a closed convex set and $x$ be a point in $S$, we define the following sets: the radial cone

$$R_S(x) := \bigcup_{t > 0} \{t^{-1}(S - x)\},$$

the tangent cone

$$T_S(x) := \text{cl}[R_S(x)],$$

(that is, the closure of the radial cone), and the normal cone

$$N_S(x) := \{x^* \in \mathcal{X}^* | \langle x^*, z - x \rangle \leq 0, \forall z \in S\}.$$  

**Definition 2.4.** [17, Definition 12.1] We say a set-valued mapping $T : \mathcal{X} \rightrightarrows \mathcal{X}$ is monotone if it has the property that

$$\langle v_1 - v_0, x_1 - x_0 \rangle \geq 0 \text{ whenever } v_0 \in T(x_0), v_1 \in T(x_1),$$

and strictly monotone if this inequality is strict when $x_0 \neq x_1$.

**Lemma 2.5.** (i) For any proper convex function $f : \mathcal{X} \to \mathbb{R}$, the mapping $\partial f : \mathcal{X} \rightrightarrows \mathcal{X}^*$ is monotone.

(ii) For a nonempty closed convex set $C \subset \mathcal{X}$, the mapping $N_C : \mathcal{X} \rightrightarrows \mathcal{X}^*$ is monotone.

**Proof.** See [17, Theorem 12.17, Corollary 12.18]. \qed

**Definition 2.6.** We say $f : \mathcal{X} \to (-\infty, +\infty]$ is $LC^1$ with constant $L$, if the gradient of $f$ is Lipschitz continuous with constant $L$, i.e., there exists $L \geq 0$ such that

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \forall x, y \in \mathcal{X}.$$  

For an operator $M : \mathcal{X} \to \mathcal{X}$, we denote $M > 0$ if for any $x(\neq 0) \in \mathcal{X}$, we have $\langle x, Mx \rangle > 0$. 

3. Inexact subgradient-based convex approximations method. In this section, we shall introduce the inexact subgradient-based convex approximations method (ISCA) for solving DCP. Before that, we first need to discuss the optimal conditions of problem (1).

3.1. Optimality conditions. We assume \( f_1, f_2, g_i, h_i \) (\( i = 1, 2, \ldots, p \)) are locally Lipschitz continuous and Clarke regular. Let \( G_i(x) = g_i(x) - h_i(x), i = 1, 2, \ldots, p \)

and

\[
G(x) = (G_1(x), \ldots, G_p(x))^T,
\]

the problem (1) is rewritten as

\[
\min_{x} f(x) \quad \text{s.t.} \quad x \in \Phi := \{x \in K | G(x) \leq 0\}.
\]

By [18, Proposition 7.3.5 and 7.4.3], Clarke’s subdifferentials of \( f \) and \( G_i(x) \) can be computed as

\[
\partial f(x) = \nabla f_0(x) + \partial f_1(x) - \partial f_2(x)
\]

and

\[
\partial G_i(x) = \partial g_i(x) - \partial h_i(x), \quad i = 1, \ldots, p.
\]

Here, we denote

\[
\tilde{\partial}g(x) = \partial g_1(x) \times \cdots \times \partial g_p(x),
\]

\[
\tilde{\partial}h(x) = \partial h_1(x) \times \cdots \times \partial h_p(x),
\]

\[
\tilde{\partial}G(x) = \partial G_1(x) \times \cdots \times \partial G_p(x).
\]

Then \( \tilde{\partial}G(x) = \tilde{\partial}g(x) - \tilde{\partial}h(x) \). When \( p = 1 \), we have \( \tilde{\partial}G(x) = \partial G(x) \).

Definition 3.1. Let \( \bar{x} \) be a feasible point of (1), i.e. \( \bar{x} \in \Phi \). If the condition

\[
0 \in \partial f(\bar{x}) + \tilde{\partial}G(\bar{x}) N_{R^p}(G(\bar{x})) + N_K(\bar{x})
\]

holds, we call \( \bar{x} \) a stationary point of DCP.

For a point \( y \in K \) and \( V = (v_1, \ldots, v_p)^T \in \tilde{\partial}h(y) \), we define

\[
V y := ((v_1, y), \ldots, (v_p, y))^T,
\]

\[
G_{y,V}(x) := g(x) - [h(y) + V(x - y)],
\]

and a set

\[
\Phi_y(V) := \{x \in K | G_{y,V}(x) \leq 0\}.
\]

For any \( v \in \partial f_2(y) \), define

\[
f_{y,v}(x) := \langle \nabla f_0(y), x \rangle + f_1(x) - [f_2(y) + \langle v, x - y \rangle].
\]

Since \( h(x) \) is convex, we have

\[
g(x) - h(x) \leq g(x) - [h(y) + V(x - y)], \quad \forall x \in K,
\]

which implies \( \Phi_y(V) \subseteq \Phi \). Moreover, we have \( f_{y,v}(x) \) and \( \Phi_y(V) \) are convex.

Denote the problem

\[
\min \{f_{y,v}(x) : x \in \Phi_y(V)\}
\]

by CP\((y, V, v)\). Then for any \( y \in K, V \in \tilde{\partial}h(y) \) and \( v \in \partial f_2(y) \), CP\((y, V, v)\) is a convex conservative approximation of problem (1).
Lemma 3.2. Consider a point $\bar{x} \in K$, for any $V \in \partial h(\bar{x})$ and $v \in \partial f_2(\bar{x})$, if the relative interior of $\Phi_{\bar{x}}(V)$ is nonempty and $\bar{x} \in \text{argmin}\text{CP}(\bar{x},V,v)$, then $\bar{x}$ is a stationary point of DCP.

Proof. Note that CP($\bar{x}, V, v$) is a convex optimization problem and the generalized Slater condition holds, then $\bar{x}$ is an optimal solution of CP($\bar{x}, V, v$) if and only if $\bar{x} \in \Phi_{\bar{x}}(V)$ and

$$
\begin{align*}
\begin{aligned}
0 & \in \nabla f_0(\bar{x}) + \partial f_1(\bar{x}) - v + (\partial g(\bar{x}) - V)^T \lambda + N_K(\bar{x}), \\
\lambda & \in N_{\mathbb{R}^n} (g(\bar{x}) - h(\bar{x})).
\end{aligned}
\end{align*}
$$

By Definition 3.1, $\bar{x}$ is a stationary point.

Lemma 3.2 suggests a simple iterative approach to solve DCP by computing

$$
x_{k+1} \in \text{argmin}\text{CP}(x_k, V_k, v_k),
$$

where $V_k \in \partial h(x_k)$ and $v_k \in \partial f_2(x_k)$, which is also the basic idea of sequential convex approximation method used in [10]. The result in Lemma 3.2 can also be characterized by generalized equation with monotone operators. For any $y \in K, V \in \partial h(y)$ and $v \in \partial f_2(y)$, define a set-valued mapping

$$
T_{y,V,v}(x) := \nabla f_0(y) + \partial f_1(x) - v + (\partial g(x) - V)^T N_{\mathbb{R}^n} (G_{y,V}(x)) + N_K(x).
$$

Lemma 3.3. For any $y \in K, V \in \partial h(y)$ and $v \in \partial f_2(y)$, the mapping $T_{y,V,v}(x)$ is monotone. If there exist a point $\bar{x} \in \mathbb{R}^n$ and $\bar{V} \in \partial h(\bar{x}), \bar{v} \in \partial f_2(\bar{x})$ such that $0 \in T_{\bar{x},\bar{V},\bar{v}}(\bar{x})$, then $\bar{x}$ is a stationary point of DCP.

Proof. For any $x^1, x^2 \in \Phi_{\bar{x}}(V)$ and $w^i \in T_{y,V,v}(x^i), i = 1, 2$, there exist

$$
\xi^i \in \partial f_1(x^i), W^i \in \partial g(x^i), \lambda^i \in N_{\mathbb{R}^n} (G_{y,V}(x^i)), \eta^i \in N_K(x^i), i = 1, 2,
$$

such that

$$
w^i = \nabla f_0(y) + \xi^i - v + [W^i - V]^T \lambda^i + \eta^i, \quad i = 1, 2.
$$

By Lemma 2.5, we can easily obtain

$$
\langle x^1 - x^2, \xi^1 - \xi^2 \rangle \geq 0, \quad \langle x^1 - x^2, \eta^1 - \eta^2 \rangle \geq 0,
$$

and

$$
[G_{y,V}(x^1) - G_{y,V}(x^2)]^T (\lambda^1 - \lambda^2) \geq 0.
$$

Note that $\lambda^1, \lambda^2 \geq 0$, then

$$
\begin{align*}
\langle x^1 - x^2, w^1 - w^2 \rangle &= \langle x^1 - x^2, (\xi^1 - \xi^2) + (\eta^1 - \eta^2) + [W^1 - V]^T \lambda^1 - [W^2 - V]^T \lambda^2 \rangle \\
&\geq \langle x^1 - x^2, [W^1 - V]^T \lambda^1 - [W^2 - V]^T \lambda^2 \rangle \\
&= \langle W^1(x^1 - x^2)^T \lambda^1 + [W^2(x^2 - x^1)]^T \lambda^2 - [V(x^1 - x^2)]^T (\lambda^1 - \lambda^2) \rangle \\
&\geq \langle g(x^1) - g(x^2)]^T \lambda^1 + [g(x^2) - g(x^1)]^T \lambda^2 - [V(x^1 - x^2)]^T (\lambda^1 - \lambda^2) \rangle \\
&= \langle g(x^1) - g(x^2) - V(x^1 - x^2)^T (\lambda^1 - \lambda^2) \rangle \\
&= \langle G_{y,V}(x^1) - G_{y,V}(x^2)\rangle^T (\lambda^1 - \lambda^2) \geq 0,
\end{align*}
$$

which implies the mapping $T_{y,V,v}$ is monotone.

If $0 \in T_{\bar{x},\bar{V},\bar{v}}(\bar{x})$, it is clear that $\bar{x}$ is a stationary point of DCP.
3.2. Algorithm. The inexact subgradient-based convex approximations algorithm can be stated as follows.

**Algorithm ISCA**

**Step 0.** Select constants $\sigma \in (0, 1)$ and $\rho \in (0, 1)$. Given $x^0 \in \Phi$, set $k = 0$.  

**Step 1.** Choose $M^k > 0, v^k \in \partial f_2(x^k), V^k \in \partial h(x^k)$ and $\epsilon_k \in (0, 1)$.  

**Step 2.** Obtain $d^k$ by solving the following convex problem

$$
\begin{align*}
\min_{d} & \quad Q_k(d) \\
\text{s.t.} & \quad g(x^k + d) - h(x^k) - V^k d \leq 0,
\end{align*}
$$

approximately such that

$$
\text{dist}(0, S_k(x^k + d^k)) \leq \epsilon_k \|d^k\|,  
$$

where

$$
Q_k(d) := \langle \nabla f_0(x^k), d \rangle + \frac{1}{2} \langle d, M^k d \rangle + f_1(x^k + d) - f_1(x^k) - \langle v^k, d \rangle
$$

and

$$
S_k(x) := T_{x^k, v^k}(x) + M^k(x - x^k).
$$

**Step 3.** Choose $\alpha^0_k > 0$. Let $l_k$ be the smallest nonnegative integer $l$ satisfying

$$
f(x^k + \alpha^0_k \rho^l d^k) \leq f(x^k) + \sigma \alpha^0_k \rho^l \Delta^k,
$$

where

$$
\Delta^k = \langle \nabla f_0(x^k), d^k \rangle + f_1(x^k + d^k) - f_1(x^k) - \langle v^k, d^k \rangle.
$$

Set

$$
\alpha_k = \alpha^0_k \rho^l, \quad x^{k+1} = x^k + \alpha_k d^k.
$$

**Step 4.** If $x^{k+1} = x^k$, then stop. Otherwise, replace $k$ by $k + 1$ and go to **Step 1.**

A similar algorithm was introduced in [8] for solving the unconstrained problem

$$
\begin{align*}
\min_{x} & \quad f(x) := f_0(x) + f_1(x) - f_2(x) \\
\text{s.t.} & \quad x \in \mathcal{X}.
\end{align*}
$$

In this paper, we extend the algorithm to a more general form. Furthermore, from the computational point of view, the major cost of Algorithm ISCA is solving the subproblem (20) in Step 2. The cost of solving problem (20) exactly at each iteration is prohibitive, which motivates us to modify the algorithm to an inexact version. As shown in Step 2, at each iteration we only approximately solve the subproblem.

In Step 3, we choose $\rho^l$ by using Armijo line search rule, which can be replaced by various line search rules if $f_1(x)$ and $f_2(x)$ are smooth. In the following two lemmas, we shall show that the Armijo rule is well defined in our algorithm.

**Lemma 3.4.** Let the sequences $\{x^k\}$ and $\{d^k\}$ be generated by Algorithm ISCA. Then for $\alpha \in (0, 1], k \geq 1$, we have

$$
f(x^k + \alpha d^k) - f(x^k) \leq \alpha \Delta^k + o(\alpha)
$$

and

$$
\Delta^k \leq -\langle d^k, M^k d^k \rangle + \epsilon_k \|d^k\|^2.
$$
Proof. From the convexity of $f_1$ and $f_2$, we obtain
\[
f(x^k + \alpha d^k) - f(x^k) = \alpha \langle \nabla f_0(x^k), d^k \rangle + \alpha f_1(x^k + \alpha d^k) - f_1(x^k) - f_2(x^k + \alpha d^k) + f_2(x^k)
\]
\[
\leq \alpha \langle \nabla f_0(x^k), d^k \rangle + \alpha (f_1(x^k + d^k) - f_1(x^k)) - f_2(x^k + \alpha d^k) + f_2(x^k)
\]
\[
\leq \alpha \langle \nabla f_0(x^k), d^k \rangle + \alpha (f_1(x^k + d^k) - f_1(x^k)) - f_2(x^k + \alpha d^k) + f_2(x^k)
\]
\[
= \alpha \Delta^k + o(\alpha),
\]
which proves (24).

By the definition of $S_k(x)$ and (21), there exists
\[
w^k \in S_k(x^k + d^k) = T_{x^k, V^k,v^k}(x^k + d^k) + M^k d^k
\]
such that $\|w^k\| \leq \varepsilon_k\|d^k\|$. Recall the definition of $T_{x^k, V^k,v^k}(x^k + d^k)$ in (19), there exist $\xi^k \in \partial f_1(x^k + d^k), W^k \in \partial g(x^k + d^k), \lambda^k \in \mathbb{R}^N (G_{x^k, V^k}(x^k + d^k))$ and $\eta^k \in N_K(x^k + d^k)$ such that
\[
w^k = \nabla f_0(x^k) + \xi^k - v^k + [W^k - V^k]^T \lambda^k + \eta^k + M^k d^k.
\]
By $\eta^k \in N_K(x^k + d^k)$ and $x^k \in K$, we obtain
\[
\langle \eta^k, d^k \rangle \geq 0. \tag{26}
\]
By $\lambda^k \in \mathbb{R}^p (G_{x^k, V^k}(x^k + d^k))$, we get
\[
\lambda^k_i \geq 0, \lambda^k_i (G_{x^k, V^k})_{i}(x^k + d^k) = 0, \ i = 1, \ldots, p,
\]
which together with $W^k \in \partial g(x^k + d^k)$ and the fact $G(x^k) \leq 0$ imply
\[
\langle [W^k - V^k]^T \lambda^k, d^k \rangle \geq [g(x^k + d^k) - g(x^k) - V^k d^k]^T \lambda^k
\]
\[
= [g(x^k + d^k) - h(x^k) - V^k d^k - (g(x^k) - h(x^k))]^T \lambda^k \tag{27}
\]
\[
= [G_{x^k, V^k}(x^k + d^k) - G(x^k)]^T \lambda^k \geq 0.
\]
Then, by (26) and (27) we have
\[
\langle [W^k - V^k]^T \lambda^k + \eta^k, d^k \rangle \geq 0.
\]
Therefore,
\[
\Delta^k = \langle \nabla f_0(x^k), d^k \rangle + f_1(x^k + d^k) - f_1(x^k) - \langle v^k, d^k \rangle
\]
\[
\leq \langle \nabla f_0(x^k), d^k \rangle + \langle \xi^k, d^k \rangle - \langle v^k, d^k \rangle
\]
\[
= \langle w^k, d^k \rangle - \langle [W^k - V^k]^T \lambda^k + \eta^k, d^k \rangle - \langle M^k d^k, d^k \rangle
\]
\[
\leq \varepsilon_k \|d^k\|^2 - \langle d^k, M^k d^k \rangle.
\]
The proof is complete. \(\square\)

Assumption 1. For all $k \geq 0$, $\underline{\nu}\|d\|^2 \leq \langle d, M^k d \rangle \leq \bar{\nu}\|d\|^2$ for any $d \in K$, where $0 < \underline{\nu} \leq \bar{\nu} < +\infty$.

Note that Assumption 1 can be easily satisfied by choosing $M^k$ appropriately.
**Lemma 3.5.** Let the sequences \( \{x^k\} \) and \( \{d^k\} \) be generated by Algorithm ISCA. Suppose that Assumption 1 holds and \( f_0 \) is LC\( ^2 \) with constant \( L \geq 0 \). For each \( k \geq 0 \), if \( 2\varepsilon_k \leq \nu \), then the descent condition
\[
f(x^k + \alpha d^k) - f(x^k) \leq \sigma \alpha \Delta^k \tag{28}
\]
is satisfied for any \( \sigma \in (0, 1) \) whenever \( 0 \leq \alpha \leq \min\{1, \nu(1 - \sigma)/L\} \).

**Proof.** By Lemma 3.4, we obtain
\[
-\Delta^k + \varepsilon_k \|d^k\|^2 \geq \langle d^k, M^k d^k \rangle \geq \nu \|d^k\|^2,
\]
and thus by \( 2\varepsilon_k \leq \nu \),
\[
\|d^k\|^2 \leq \frac{-2\Delta^k}{\nu}.
\]
For any \( \alpha \in (0, 1] \), we have
\[
f(x^k + \alpha d^k) - f(x^k)
\]
\[
= f_0(x^k + \alpha d^k) - f_0(x^k) + f_1(x^k + \alpha d^k) - f_1(x^k) - [f_2(x^k + \alpha d^k) - f_2(x^k)]
\]
\[
= \alpha \langle \nabla f_0(x^k), d^k \rangle + \int_0^1 \langle \nabla f_0(x^k + t\alpha d^k) - \nabla f_0(x^k), \alpha d^k \rangle dt
\]
\[
+ f_1(x^k + \alpha d^k) - f_1(x^k) - [f_2(x^k + \alpha d^k) - f_2(x^k)]
\]
\[
\leq \alpha \langle \nabla f_0(x^k), d^k \rangle + \alpha \int_0^1 \langle \nabla f_0(x^k + t\alpha d^k) - \nabla f_0(x^k), d^k \rangle dt
\]
\[
+ \alpha [f_1(x^k + \alpha d^k) - f_1(x^k)] - \alpha \langle v^k, d^k \rangle
\]
\[
= \alpha \Delta^k + \alpha \int_0^1 \langle \nabla f_0(x^k + t\alpha d^k) - \nabla f_0(x^k), d^k \rangle dt
\]
\[
\leq \alpha \Delta^k + \frac{L}{2} \alpha^2 \|d^k\|^2
\]
\[
\leq \alpha \Delta^k (1 - \frac{L\alpha}{\nu}).
\]
Therefore, for any \( \sigma \in (0, 1) \), whenever \( 0 \leq \alpha \leq \min\{1, \nu(1 - \sigma)/L\} \), we have
\[
f(x^k + \alpha d^k) - f(x^k) \leq \sigma \alpha \Delta^k.
\]

3.3. **Convergence analysis.** Under the conditions in Lemma 3.5, if there exists an integer \( k \geq 0 \) such that \( x^{k+1} = x^k \) in Algorithm ISCA, we will show \( x^k \) is a stationary point of problem (1). Otherwise, we will show any accumulation of the infinite sequence \( \{x^k\} \) is a stationary point.

**Theorem 3.6.** Let the sequences \( \{x^k\} \) and \( \{d^k\} \) be generated by Algorithm ISCA. Suppose that Assumption 1 holds and \( f_0 \) is LC\( ^2 \) with constant \( L \geq 0 \). For some integer \( k \geq 0 \), if the relative interior of \( \Phi_{x^k}(V^k) \) is nonempty and \( x^{k+1} = x^k \), then \( x^k \) is a stationary point of problem (1).

**Proof.** Since \( d^k = 0 \), by (21) we have \( 0 \in S_k(x^k) \), thus \( 0 \in T_{x^k, V^k, v^k}(x^k) \). By Lemma 3.3, we obtain \( x^k \) is a stationary point of problem (1).

**Theorem 3.7.** Let the sequences \( \{x^k\} \) and \( \{d^k\} \) be generated by Algorithm ISCA. Suppose that Assumption 1 holds and \( f_0 \) is LC\( ^2 \) with constant \( L \geq 0 \). For each \( k \geq 0 \), suppose that \( 2\varepsilon_k \leq \nu \) and \( \inf_k \alpha^0_k > 0 \), the following results hold:
(i). For each integer $k \geq 0$, $f(x^{k+1}) - f(x^k) \leq \sigma \alpha_k \Delta^k \leq 0$.
(ii). If $\{x^k\}$ is a convergent subsequence of $\{x^k\}$, then $\lim_{j \to +\infty} d^{k_j} = 0$.
(iii). Suppose that $(x, V, v)$ is an accumulation point of $\{(x^k, V^k, v^k)\}$ and the relative interior of $\Phi_x(V)$ is nonempty, then $x$ is a stationary point of problem (1).

Proof. In this theorem, we consider the case that $x^{k+1} \neq x^k$ for all $k \geq 0$.

(i). From Lemma 3.4 and Assumption 1, we have
\[
\Delta^k \leq \frac{1}{2} \nu \|d^k\|^2 \leq 0.
\]

Then, by line search condition (22), we obtain
\[
f(x^{k+1}) - f(x^k) \leq \sigma \alpha_k \Delta^k \leq 0.
\]

(ii). Let $\bar{x} := \lim_{j \to +\infty} x^{k_j}$. Since $f$ is continuous, $\lim_{j \to +\infty} f(x^{k_j}) = f(\bar{x})$. Note that $\{f(x^k)\}$ is a non-increasing sequence, which implies $\lim_{j \to +\infty} f(x^{k_j}) = f(\bar{x})$. Then from (i) we have
\[
\alpha_k \Delta^k = 0.
\]

Suppose that $d^{k_j} \to 0$ when $j \to +\infty$. By passing to a subsequence if necessary, we can assume that, for some $\delta > 0$, $\|d^{k_j}\| \geq \delta$ for all $j \geq 0$. Thus,
\[
\Delta^{k_j} \leq \frac{1}{2} \nu \|d^{k_j}\|^2 \leq \frac{1}{2} \mu \delta^2 < 0.
\]

By (29) we have $\alpha_k \to 0$. Together with $\alpha_k = \alpha^0_k \rho^{k_j}$ and $\inf_k \alpha^0_k > 0$, there exists $k > 0$ such that for any $k_j > k$,
\[
\alpha_{k_j} = \alpha^0_k \rho^{k_j}, \quad \alpha_k \leq \rho.
\]

Furthermore, since $\alpha_k$ is chosen by the Armijo rule, it implies that
\[
f(x^{k_j} + (\alpha_k \rho^{k_j})^{k_j} - f(x^{k_j}) > (\alpha_k / \rho) \Delta^{k_j}, \quad \forall k_j > k.
\]

Thus,
\[
s \Delta^{k_j} = \frac{f(x^{k_j} + (\alpha_k \rho^{k_j})^{k_j}) - f(x^{k_j})}{\alpha_k \rho^{k_j}} \leq \frac{f_0(x^{k_j} + (\alpha_k \rho^{k_j})^{k_j}) - f_0(x^{k_j})}{\alpha_k \rho^{k_j}} + f_1(x^{k_j} + d^{k_j}) - f_1(x^{k_j} - (v^{k_j}, d^{k_j})
\]
\[
= \frac{f_0(x^{k_j} + (\alpha_k \rho^{k_j})^{k_j}) - f_0(x^{k_j})}{\alpha_k \rho^{k_j}} - \langle \nabla f_0(x^{k_j}), d^{k_j} \rangle + \Delta^{k_j}.
\]

It implies that
\[
\frac{f_0(x^{k_j} + (\alpha_k \rho^{k_j})^{k_j}) - f_0(x^{k_j})}{\alpha_k \rho^{k_j}} - \langle \nabla f_0(x^{k_j}), d^{k_j} \rangle > (\sigma - 1) \Delta^{k_j} \geq \frac{1}{2} (1 - \sigma) \nu \|d^{k_j}\|^2,
\]

or equivalently
\[
\frac{f_0(x^{k_j} + (\alpha_k \rho^{k_j})^{k_j}) - f_0(x^{k_j})}{\alpha_k \|d^{k_j}\| \rho^{k_j}} - \langle \nabla f_0(x^{k_j}), d^{k_j} \rangle \geq \frac{1}{2} (1 - \sigma) \nu \|d^{k_j}\| > 0.
\]

From the fact
\[-\alpha_k \Delta^{k_j} \geq \delta \nu \alpha_k \|d^{k_j}\| \geq 0,
\]
we have \(\alpha_k \|d^k\|/\rho \to 0\) as \(j \to +\infty\). Note that there exists \(\tilde{d}\) with \(\|\tilde{d}\| = 1\), by further passing to a subsequence if necessary, such that
\[
\lim_{j \to +\infty} \frac{d^k_j}{\|d^k_j\|} = \tilde{d}.
\]
By taking \(j \to +\infty\) in (30) we have
\[
0 = \langle \nabla f_0(\bar{x}), \tilde{d} \rangle - \langle \nabla f_0(\bar{x}), \tilde{d} \rangle > 0,
\]
which is a contradiction.

(iii). Suppose that \((\bar{x}, V, v)\) is an accumulation point of \(\{(x^k, V^k, v^k)\}\), there exists a subsequence \(\{k_j\}\) such that \((x^{k_j}, V^{k_j}, v^{k_j}) \to (\bar{x}, V, v)\) as \(k_j \to +\infty\). Since the relative interior of \(\Phi_{\bar{x}}(V)\) is nonempty, there exists \(\hat{d} \in \mathbb{R}^n\) such that
\[
\bar{x} + \hat{d} \in \text{int}K,
\]
\[
g(\bar{x} + \hat{d}) - h(\bar{x}) - V \hat{d} < 0.
\]
Then for sufficiently large \(j\), \(x^{kj} + \hat{d}\) is in the relative interior of \(\Phi_{x^{kj}}(V^{k_j})\). From the inequality
\[
dist((0, S_{k_j}(x^{kj} + d^{kj}))) \leq \varepsilon_{k_j}\|d^{kj}\|,
\]
there exists \(w^{kj} \in S_{k_j}(x^{kj} + d^{kj})\) such that \(\|w^{kj}\| \leq \varepsilon_{k_j}\|d^{kj}\|\). Thus there exist \(\xi^{kj} \in \partial f_1(x^{kj} + d^{kj}), W^{kj} \in \partial g(x^{kj} + d^{kj})\) and \(\lambda^{kj} \in \mathbb{R}^p\) such that
\[
\begin{aligned}
  w^{kj} &\in \nabla f_0(x^{kj}) + \xi^{kj} - v^{kj} + [W^{kj} - V^{k_j}]^T \lambda^{kj} + N_K(x^{kj} + d^{kj}) + M^{k_j}d^{kj}, \\
  0 &\leq \lambda^{kj}g(x^{kj} + d^{kj}) - h(x^{kj}) - V^{k_j}d^{kj} \leq 0.
\end{aligned}
\]
Since \(x^{kj} \to \bar{x}\) and \(d^{kj} \to 0\), we have \(x^{kj} + d^{kj} \to \bar{x}\) and \(w^{kj} \to 0\). From the outer semicontinuity of \(\partial f_1, \partial f_2, \partial g\) and \(\partial h\) at \(\bar{x}\), for sufficiently large \(j_0\), the sequences \(\{\xi^{k_j}\}_{j \geq j_0}, \{v^{k_j}\}_{j \geq j_0}, \{W^{k_j}\}_{j \geq j_0} \) and \(\{V^{k_j}\}_{j \geq j_0}\) are bounded. By taking further subsequences respectively, if necessary, there exist \(\xi \in \partial f_1(\bar{x}), v \in \partial f_2(\bar{x}), W \in \partial g(\bar{x})\) and \(V \in \partial h(\bar{x})\) such that
\[
\xi^{k_j} \to \xi, \ v^{k_j} \to v, \ W^{k_j} \to W, \ V^{k_j} \to V.
\]

Next we shall prove that \(\{\lambda^{k_j}\}\) is also bounded. Suppose not. Passing to a subsequence if necessary, we assume that \(\|\lambda^{k_j}\| \to +\infty\). Let
\[
\tilde{\lambda}^{k_j} = \lambda^{k_j}/\|\lambda^{k_j}\|.
\]
Passing to a subsequence if necessary, we assume \(\tilde{\lambda}^{k_j} \to \tilde{\lambda}\), then
\[
\tilde{\lambda}^{k_j} \in N_{\mathbb{R}^p}(g(x^{kj} + d^{kj}) - h(x^{kj}) - V^{k_j}d^{kj})
\]
and
\[
\tilde{\lambda} \in N_{\mathbb{R}^p}(G(\bar{x})), \ |\tilde{\lambda}| = 1.
\]
Divide both sides of the first formula in (31) by \(\|\lambda^{k_j}\|\),
\[
\begin{aligned}
\frac{w^{k_j}}{\|\lambda^{k_j}\|} &\in \frac{\nabla f_0(x^{kj}) + \xi^{kj} - v^{kj} + M^{k_j}d^{kj}}{\|\lambda^{k_j}\|} + [W^{kj} - V^{k_j}]^T \tilde{\lambda}^{k_j} + N_K(x^{kj} + d^{kj}), \\
0 &\leq [W^{kj} - V^{k_j}]^T \tilde{\lambda} + N_K(x^{kj} + d^{kj}),
\end{aligned}
\]
and let \(j \to +\infty\) we have
\[
0 \in [W - V]^T \tilde{\lambda} + N_K(\bar{x}). \tag{32}
\]
From $\bar{x} + \hat{d} \in \text{int}K$ we have
\[ 0 < \langle \hat{d}, [W - V]^T \hat{\lambda} \rangle. \] (33)

Define
\[ I(\bar{x}) := \{ i | G_i(\bar{x}) = 0, i = 1, 2, \ldots, p \}. \]

From $\tilde{\lambda} \in N_{\mathbb{R}^p}(G(\bar{x}))$, for $i = 1, 2, \ldots, p$, $\tilde{\lambda}_i \geq 0$, and if $i \notin I(\bar{x})$, $\tilde{\lambda}_i = 0$. Then by (33),
\[ \sum_{i \in I(\bar{x})} \tilde{\lambda}_i [\langle w_i, \hat{d} \rangle - \langle v_i, \hat{d} \rangle] > 0. \] (34)

From $g(\bar{x} + \hat{d}) - h(\bar{x}) - V \hat{d} < 0$, for $i \in I(\bar{x})$,
\[ 0 > g_i(\bar{x} + \hat{d}) - h_i(\bar{x}) - \langle v_i, \hat{d} \rangle \]
\[ \geq g_i(\bar{x}) + \langle w_i, \hat{d} \rangle - h_i(\bar{x}) - \langle v_i, \hat{d} \rangle \]
\[ = G_i(\bar{x}) + \langle w_i, \hat{d} \rangle - \langle v_i, \hat{d} \rangle \]
\[ = \langle w_i, \hat{d} \rangle - \langle v_i, \hat{d} \rangle. \]

Thus we obtain a contradiction, which implies $\{ \lambda^{kj} \}$ is bounded. By taking further subsequence if necessary, there exists $\lambda \in N_{\mathbb{R}^p}(G(\bar{x}))$ such that $\lambda^{kj} \rightarrow \lambda$. By taking $j \rightarrow +\infty$ in (31), we have
\[ \begin{cases} 0 \in \nabla f_0(\bar{x}) + \xi - v + [W - V]^T \lambda + N_K(\bar{x}), \\ 0 \leq \lambda \perp G(\bar{x}) \leq 0, \end{cases} \]
which implies $\bar{x}$ is a stationary point of problem (1).

3.4. Implementation issues. In this subsection, we address several practical issues in the implementation of algorithm ISCA.

(i.) The choice of the initial point $x^0 \in \Phi$. First we find an $\bar{x} \in K$, and choose $v \in \partial f_2(\bar{x}), V \in \partial h(\bar{x})$, then obtain $\hat{d}$ which is an optimal point of the following convex problem:
\[ \begin{align*}
\min & \quad \langle \nabla f_0(\bar{x}), d \rangle + f_1(\bar{x} + d) - f_1(\bar{x}) - \langle v, d \rangle \\
\text{s.t.} & \quad g(\bar{x} + d) - h(\bar{x}) - Vd \leq 0, \\
& \quad \bar{x} + d \in K.
\end{align*} \] (35)

Then, the point $x_0 = \bar{x} + \hat{d} \in \Phi$ may be a good choice.

(ii.) The choice of the algorithm for solving the subproblems (20). The performance of our algorithm heavily relies on the algorithm for solving the convex subproblems of the form (20). For various applications, the subproblems will be greatly different. For instance, when solving joint chance constrained programs by the Monte Carlo method, we will see the constraint function of the subproblem is a sum of a large number of convex functions. See another example in [8] for a structured low rank matrix optimization problem, the subproblem is a least square positive semidefinite constrained program. Thus, it is impossible for us to give a method which is suitable for various kinds of subproblems. In the numerical experiments, we apply a standard MATLAB solver fmincon for the subproblems.

4. Numerical experiments. In this section, we consider two kinds of numerical problems, and use them to illustrate the performances of our method.
4.1. A joint chance constrained problem. We consider a stochastic program constructed in [10]:

$$\begin{align*}
\max & \quad \|x\|_1 \\
\text{s.t.} & \quad \Pr \left\{ \sum_{j=1}^{n} \xi_{ij}^2 x_j^2 - b \leq 0, \quad i = 1, \ldots, m \right\} \geq 1 - \alpha, \\
\end{align*}$$

(36)

where $\xi_{ij}, i = 1, \ldots, m$ and $j = 1, \ldots, n$ are independent and identically distributed standard norm random variables, $b$ is a number and

$$\|x\|_1 := \sum_{i=1}^{n} |x_i|$$

is the $l_1$ norm of $x$. Define

$$c_i(x, \xi) := \sum_{j=1}^{n} \xi_{ij}^2 x_j^2 - b,$$

and

$$c(x, \xi) := \max \{c_1(x, \xi), \ldots, c_m(x, \xi)\}.$$ 

The problem (36) can be reformulated as

$$\begin{align*}
\min & \quad -\sum_{j=1}^{n} x_j \\
\text{s.t.} & \quad \Pr\{c(x, \xi) \leq 0\} \geq 1 - \alpha, \\
& \quad x \geq 0.
\end{align*}$$

(37)

Note that problem (37) is a chance constrained problem as discussed in subsection 1.3. For any $\varepsilon > 0$, let

$$g_1(x, \varepsilon) = \frac{1}{\varepsilon} \mathbb{E}[c(x, \xi) + \varepsilon]_+$$

and

$$g_2(x, \varepsilon) = \frac{1}{\varepsilon} \mathbb{E}[c(x, \xi)]_+,$$

then, as shown in subsection 1.3, the constraint $\Pr\{c(x, \xi) \leq 0\} \geq 1 - \alpha$ can be approximated by $g_1(x, \varepsilon) - g_2(x, \varepsilon) \leq \alpha$ for a small $\varepsilon$. And a good approximation of problem (37) is given by

$$\begin{align*}
\min & \quad -\sum_{j=1}^{n} x_j \\
\text{s.t.} & \quad g_1(x, \varepsilon) - g_2(x, \varepsilon) \leq \alpha, \\
& \quad x \geq 0.
\end{align*}$$

(P_\varepsilon)

In [10] the authors proved that, under mild assumptions, the KKT point of problem (P_\varepsilon) converges to the KKT point of problem (37) as $\varepsilon \to 0$. Note that problem (P_\varepsilon) is a DC program, thus our method can be applied to approximately solve problem (37).

**Lemma 4.1** (Lemma 7.2, [19]). The optimal solution $x^*$ of problem (37) is

$$x_1^* = x_2^* = \cdots = x_n^* = \left[ \frac{b}{F_{\chi_n^2}^{-1}\left((1 - \alpha)^{\frac{1}{m}}\right)} \right]^{\frac{1}{2}},$$

where $F_{\chi_n^2}^{-1}$ is the inverse chi-squared distribution function with $n$ degrees of freedom.
To implement Algorithm ISCA for solving problem \((P_\varepsilon)\), we need to know the closed form expressions of \(g_1(x, \varepsilon)\) and \(g_2(x, \varepsilon)\) which are generally difficult to obtain. However, we can use Monte Carlo method to get the approximations of those expressions. Let \(\xi_1, \ldots, \xi_N\) be an independent and identically distributed (i.i.d.) sample of \(\xi\). Let

\[
\bar{g}_1(x, \varepsilon) = \frac{1}{N\varepsilon} \sum_{l=1}^{N} [c(x, \xi_l) + \varepsilon],
\]

and

\[
\bar{g}_2(x, \varepsilon) = \frac{1}{N\varepsilon} \sum_{l=1}^{N} [c(x, \xi_l)].
\]

Then \(\bar{g}_1(x, \varepsilon)\) and \(\bar{g}_2(x, \varepsilon)\) are the sample average approximations of \(g_1(x, \varepsilon)\) and \(g_2(x, \varepsilon)\), respectively.

The numerical experiment is carried out in MATLAB 7.10 running on a PC Intel Core i5 CPU (3.20 GHz) and 4 GB RAM. In the following test problems, we set \(\alpha = 0.1\). We terminate our algorithm if

\[
\frac{\|d^k\|}{\|x^k\|} \leq 10^{-5}.
\]

**Example 1.** In problem (37), let \(n = 3, m = 2, b = 9, N = 10,000\). By Lemma 4.1, the optimal solution is \((1.08, 1.08, 1.08)^T\), the optimal value is about -3.23.

**Example 2.** In problem (37), let \(n = 10, m = 10, b = 100, N = 10,000\). By Lemma 4.1, the optimal solution is \((2.08, 2.08, \cdots, 2.08)^T\), the optimal value is about -20.82.

**Figure 2.** Numerical results for Example 1 and 2

The numerical results are shown in Figure 2. Take Example 2 for instance, the algorithm typically requires less than 10 iterations to converge to the optimal value. When sample size \(N = 10,000\), at each iteration, a convex program is solved in 6 seconds on average.

4.2. **Homogeneous quadratic constrained quadratic optimization.** In this subsection, we consider a homogeneous quadratic optimization problem in the form of

\[
\min_{x} \quad x^* A_0 x \\
\text{s.t.} \quad x^* A_k x \geq 1, \quad k = 1, \cdots, m, \\
x \in \mathbb{F}^n,
\]

\((\text{QCQP})\)
where $A_k, k = 0, 1, \ldots, m$ are $n \times n$ real symmetric or complex Hermitian matrices, $F$ is either the field of real numbers $\mathbb{R}$ or the field of complex numbers $\mathbb{C}$, and the superscript $*$ denotes regular transpose or Hermitian transpose. The homogeneous quadratic constrained quadratic optimization problem (QCQP) arises in the transmit beamforming problem for multicasting applications [21, 13, 9] and the references therein. The QCQP problem is NP-hard [13], even when all matrices are positive semidefinite. A natural approach for solving QCQP is to use its SDP relaxation:

$$\begin{align*}
\text{(SDP)} \\
\min & \quad \text{Tr}(A_0 X) \\
\text{s.t.} & \quad \text{Tr}(A_k X) \geq 1, \quad k = 1, \ldots, m, \\
& \quad X \in \mathbb{S}_+^n,
\end{align*}$$

where $\text{Tr}(\cdot)$ denotes the trace of a matrix, and $\mathbb{S}_+^n$ denotes the cone of all (Hermitian) positive semidefinite matrices. In [9] it is shown that if the matrix $A_0$ and all but one of the $A_k, k = 1, \ldots, m$ are positive semidefinite, the ratio between the optimal value of QCQP and its SDP relaxation is upper bounded by $O(m^2)$ when $F = \mathbb{R}$, and by $O(m)$ when $F = \mathbb{C}$. Moreover, when two or more of the $A_k, k = 1, \ldots, m$ are indefinite, the ratio can be arbitrarily large.

Here, we propose a DC programming method for solving QCQP. For each $k = 0, 1, \ldots, m$, let $\lambda_k$ be the largest eigenvalue of $A_k$. Let $\rho_k := \max\{0, \lambda_k\}$ and $I$ be the identity matrix. Define

$$g_k(x) := x^*(\rho_k I - A_k)x, \quad h_k(x) := \rho_k x^* x,$$

then QCQP can be rewritten as:

$$\begin{align*}
\min & \quad h_0(x) - g_0(x) \\
\text{s.t.} & \quad g_k(x) + 1 - h_k(x) \leq 0, \quad k = 1, \ldots, m, \\
& \quad x \in \mathbb{F}^n,
\end{align*}$$

which is a DC program in the form of (1). Thus we can use Algorithm ISCA to obtain a stationary point of QCQP.

**Example 3.** [9, Example 3.7]

$$\begin{align*}
\min & \quad x_4^2 \\
\text{s.t.} & \quad x_1 x_2 + x_3^2 + x_4^2 \geq 1, \\
& \quad -x_1 x_2 + x_3^2 + x_4^2 \geq 1, \\
& \quad 0.5 x_1^2 - x_3^2 \geq 1, \\
& \quad 0.5 x_2^2 - x_3^2 \geq 1, \\
& \quad x_1, x_2, x_3, x_4 \in \mathbb{R}.
\end{align*}$$

From the last two constraints we have $|x_1 x_2| \geq 2(x_3^2 + 1)$. Meanwhile, the first two constraints imply $|x_1 x_2| \leq x_3^2 + x_4^2 - 1$. Combining these two inequalities yields $x_4^2 \geq 3$. Thus, the optimal value is 3 and the optimal solutions are $x_1 = \pm \sqrt{2}, x_2 = \pm \sqrt{2}, x_3 = 0, x_4 = \pm \sqrt{3}$.

We also use our algorithm to solve the above example in MATLAB. We choose the initial point randomly by $x_0 = \text{rand}(4, 1)$. By running the algorithm 100 replications, we see that it almost always converges to one of the optimal solutions (90 replications), even though the problem is not convex. The algorithm typically requires about 15 iterations to converge.
Example 4. [13, Example 3]

\[
\begin{align*}
\min & \quad x_1^2 + x_2^2 \\
\text{s.t.} & \quad x_2^2 \geq 1, \\
& \quad x_1^2 + Mx_1x_2 \geq 1, \\
& \quad x_1^2 - Mx_1x_2 \geq 1, \\
& \quad x_1, x_2 \in \mathbb{R},
\end{align*}
\]

where \( M > 0 \) is a constant.

The last two constraints imply \( x_1^2 \geq M|x_1x_2| + 1 \) which, together with the first constraint \( x_2^2 \geq 1 \), yield \( x_1^2 \geq (M + \sqrt{M^2 + 4})/2 \). Therefore, the optimal value is \( 1 + (M + \sqrt{M^2 + 4})^2/4 \) and the optimal solutions are

\[
x_1 = \pm(M + \sqrt{M^2 + 4})/2, x_2 = \pm 1.
\]

In the numerical experiments for this example, we find that the convergence to a global optimal solution cannot be guaranteed, but somehow when we set the initial point by \( x_0 = M \ast (1, 0)^T + \text{rand}(2, 1) \), the algorithm always converges to a global optimal solution.

Acknowledgments. We are grateful to the associate editor and one referee for their helpful comments.

REFERENCES

[1] L. T. H. An, An efficient algorithm for globally minimizing a quadratic function under convex quadratic constraints, Math. Program., 87 (2000), 401–426.
[2] L. T. H. An, L. H. Minh and P. D. Tao, Optimization based DC programming and DCA for hierarchical clustering, European J. Oper. Res., 183 (2007), 1067–1085.
[3] L. T. H. An and P. D. Tao, The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems, Ann. Oper. Res., 133 (2005), 23–46.
[4] C. Audet, P. Hansen, B. Jaumard and G. Savard, A branch and cut algorithm for nonconvex quadratically constrained quadratic programming, Math. Program., 87 (2000), 131–152.
[5] J. F. Bonnans and A. Shapiro, Perturbation Analysis of Optimization Problems, Springer, New York, 2000.
[6] D. H. Fang, C. Li and X. Q. Yang, Asymptotic closure condition and Fenchel duality for DC optimization problems in locally convex spaces, Nonlinear Anal., 75 (2012), 3672–3681.
[7] M. Fazel, Matrix Rank Minimization with Applications, PhD thesis, Stanford University, 2002.
[8] Y. Gao, Structured Low Rank Matrix Optimization Problems: A Penalty Approach, PhD thesis, National University of Singapore, 2010.
[9] S. He, Z. Luo, J. Nie and S. Zhang, Semidefinite relaxation bounds for indefinite homogeneous quadratic optimization, SIAM J. Optim., 19 (2008), 503–523.
[10] L. J. Hong, Y. Yang and L. Zhang, Sequential convex approximations to joint chance constrained programs: a Monte Carlo approach, Oper. Res., 59 (2011), 617–630.
[11] R. Horst and N. V. Tho, DC programming: Overview, J. Optim. Theory Appl., 103 (1999), 1–43.
[12] Z. Luo, W. Ma, A. So, Y. Ye and S. Zhang, Semidefinite relaxation of quadratic optimization problems, IEEE Signal Process Mag., 27 (2010), 20–34.
[13] Z. Luo, N. Sidiropoulos, P. Tseng and S. Zhang, Approximation bounds for quadratic optimization with homogeneous quadratic constraints, SIAM J. Optim., 18 (2007), 1–28.
[14] B. S. Mordukhovich, Variational Analysis and Generalized Differentiation I: Basic Theory, Springer, Berlin, 2006.
[15] A. Nemirovski and A. Shapiro, Convex approximations of chance constrained programs, SIAM J. Optim., 17 (2006), 969–996.
[16] B. Recht, M. Fazel and P. A. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, SIAM Rev., 52 (2010), 471–501.
[17] R. T. Rockafellar and R. J. B. Wets, Variational Analysis, Springer, New York, 1998.
[18] W. Schirotzek, Nonsmooth Analysis, Springer, Berlin, 2007.
[19] F. Shan, L. Zhang and X. Xiao, A smoothing function approach to joint chance-constrained programs, J. Optim. Theory Appl., 59 (2014), 181–199.
[20] A. Shapiro, D. Dentcheva and A. Ruszczyński, Lectures on Stochastic Programming: Modeling and Theory, SIAM, Philadelphia, 2009.
[21] N. Sidiropoulos, T. Davidson and Z. Luo, Transmit beamforming for physical-layer multicasting, IEEE Trans. Signal Process., 54 (2006), 2239–2251.
[22] N. V. Thoai, Reverse convex programming approach in the space of extreme criteria for optimization over efficient sets, J. Optim. Theory Appl., 147 (2010), 263–277.
[23] X. Xiao, J. Gu, L. Zhang and S. Zhang, A sequential convex program method to DC program with joint chance constraints, J. Ind. Manag. Optim., 8 (2012), 733–747.

Received September 2014; revised May 2015.

E-mail address: gjian@dlou.edu.cn
E-mail address: xtxiao@dlut.edu.cn
E-mail address: lwzhang@dlut.edu.cn