A parametric kernel function with a trigonometric barrier term for second-order cone optimization

Guangyuan Che, Xiyao Luo, Fudong Chen, Qinglin Hu and Yue Zhang

College of Fundamental Studies, Shanghai University of Engineering Science, Shanghai 201620, China

Abstract

In this paper, we generalize a primal-dual interior-point algorithm based on a parametric kernel function, which was studied by M. Achache [1], for linear optimization, to second-order cone optimization. By using Jordan algebra, the currently best known iteration bounds for large-update methods is derived, namely, $O\left(\sqrt{N} \log N \log(N/\varepsilon)\right)$.

Keywords

Interior-point methods; Second-order cone optimization; Large-update methods; Polynomial complexity

Academic Discipline and Sub-Disciplines

Mathematics; Operations Research and Control Theory

SUBJECT CLASSIFICATION

Mathematics Subject Classification 2010: 90C22, 90C51

1 INTRODUCTION

Consider the standard second-order cone optimization (SOCO) problem

$$\begin{align*}
(P) & \quad \text{min} \quad \{ c^T x : Ax = b, x \leq 0 \}, \\
(D) & \quad \text{max} \quad \{ b^T y : A^T y + s = c, s \leq 0 \},
\end{align*}$$

and its dual problem

$$\begin{align*}
(P^\ast) & \quad \text{max} \quad \{ b^T y : A^T y + s = c, s \leq 0 \}, \\
(D^\ast) & \quad \text{min} \quad \{ c^T x : Ax = b, x \leq 0 \},
\end{align*}$$

Here $c \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$ are given vectors, $A \in \mathbb{R}^{m \times n}$ is a given matrix with $\text{rank}(A) = m$, and $V = V_1 \times \ldots \times V_N \subseteq \mathbb{R}^n$ is a given second-order cone with

$$V_i := \left\{ x^{(i)} = (x^{(i)}_1, x^{(i)}_{2:m}) \in \mathbb{R} \times \mathbb{R}^{m-1} : x^{(i)}_1 \geq \| x^{(i)}_{2:m} \|_2 \right\}$$

where $x^{(i)}_{2:m} := (x^{(i)}_2, \ldots, x^{(i)}_m)$ with $m_i$ is not less than 2 for $i = 1, \ldots, N$, and $n = \sum_{i=1}^N m_i$. It is well-known that SOCO problems include linear optimization (LO) and convex quadratic optimization (CQO) problems as special cases. Furthermore, SOCO problems are the special case of semidefinite optimization (SDO) and have wide range of applications such as investment portfolio, wireless sensor network and sparse optimizations, and some other relevant areas [2, 10]. For a survey, we refer the reader to the monographs [3, 17] and the references [5, 6, 7, 11, 12, 15, 16].

Bai et al. [4] first studied the eligible kernel (and barrier) functions and developed several well useful properties of them. A unified schema of primal-dual long- and short-step interior-point algorithms (IPAs) for LO is presented. Consequently, several kernel-based primal-dual long- and short-step IPAs for various optimization and complementarity problems are seriously considered, e.g., [6, 7, 15, 17, 18].

The aim of this article is to generalize the primal-dual long-step IPAs for LO based on parametric function considered in [1] to SOCO. Nevertheless, we establish the same order of the complexity of iteration bounds for large-update methods for SOCO as the one obtained for LO in [1], namely, $O\left(\sqrt{N} \log N \log(N/\varepsilon)\right)$, by choosing $q = 1 + O(N)$.

The structure of the paper goes as follows. In Section 2, we proposed some useful results on Jordan algebras, second-order cones and the proposed parametric kernel function. Section 3 is devoted to present the generic kernel-based primal-dual long-step IPAs for SOCO. In Section 4, we establish the iteration complexity of the algorithms for long-step methods. Finally, some concluding remarks are made in Section 5.
2 PRELIMINARIES

2.1 Jordan algebras and second-order cones

We recall some concepts, properties, and results of Euclidean Jordan algebras used in this paper. More details and related results can be found in the monograph [8] and the references in [2,9,12,14].

Let

\[ x := \left( x^{(1)}, \cdots, x^{(N)} \right)^T \in V \]

with \( x^{(i)} \in \mathbb{R}^m \) for \( i = 1, \cdots, N \). Then

\[ x \circ s = \left( x^{(1)} \circ s^{(1)}, \cdots, x^{(N)} \circ s^{(N)} \right)^T. \]

The spectral decomposition of \( x^{(i)} \) is given by

\[ v^{(i)} := \lambda_{\min}(v^{(i)})c_1^{(i)} + \lambda_{\max}(v^{(i)})c_2^{(i)}, \]

where

\[ c_1^{(i)} := \frac{1}{2} \left( 1, \frac{-v^{(i)}_{2m_1}}{\| v^{(i)}_{2m_1} \|} \right)^T, \quad c_2^{(i)} := \frac{1}{2} \left( 1, \frac{v^{(i)}_{2m_1}}{\| v^{(i)}_{2m_1} \|} \right)^T \]

are the eigenvectors for the eigenvalues \( \lambda_{\min}(v^{(i)}) \) and \( \lambda_{\max}(v^{(i)}) \) of the linear transformation \( L(v^{(i)}) \) defined by

\[ L(x)y = x \circ y, \]

respectively. Let \( \frac{v^{(i)}_{2m_1}}{\| v^{(i)}_{2m_1} \|} = 0 \) only if \( v^{(i)}_{2m_1} = 0 \).

According to the spectral decomposition of \( x^{(i)} \) with \( i = 1, \cdots, N \), we define the vector valued function as follows

\[ \varphi(v) := (\varphi(v^{(1)}), \cdots, \varphi(v^{(N)}))^T, \]

with

\[ \varphi(v^{(i)}) := \varphi(\lambda_{\min}(v^{(i)}))c_1^{(i)} + \varphi(\lambda_{\max}(v^{(i)}))c_2^{(i)}, \]

where \( \varphi(t) \) is any single variable function from \( \mathbb{R} \) to \( \mathbb{R} \).

Lemma 2.1 (Lemma 3.2 in [9]) Let \( \text{int} V \) denotes the interior of \( V \), and \( x, s \in V \). Then there exists a unique \( w \in \text{int} V \) such that

\[ x = P(w)s, \]

where

\[ w = P \left( s^{-\frac{1}{2}} \right) \left( P \left( s^2 \right) x \right)^{\frac{1}{2}} = P \left( x^2 \right) \left( P \left( x^2 \right) s \right)^{-\frac{1}{2}}. \]

2.2 The parametric kernel function

Consider the following parametric kernel function

\[ \varphi(t) = \frac{t^2 - 1}{2} + \frac{q^{-1}}{\log q}, \quad q > 1, t > 0, \]

which was first considered in [1] for LO. The first three derivatives of \( \varphi(t) \) with respect to \( t \) are given by
\[ \phi'(t) = t - \frac{1}{t^2} q^{\frac{1}{t^2}}, \]  
\[ \phi''(t) = 1 + \frac{\log q + 2t}{t^2} q^{\frac{1}{t^2}}, \]  
\[ \phi'''(t) = -\frac{\log^2 q + 6t \log q + 6t^2}{t^6} q^{\frac{1}{t^2}}. \]  

For the analysis of the algorithm, we recall some technical results in [1,4] without proofs.

**Lemma 2.2** Let \( \tilde{n} : [0, \infty) \to [1, \infty) \) be the inverse function of \( \phi(t) \) for \( t \geq 1 \). Then
\[ \sqrt{1+2s} \leq \tilde{n}(s) \leq 1+\sqrt{2s}. \]

The following lemma provides the lower bounds of \( \rho(s) \), which is the inverse function of \( -\frac{1}{2} \phi'(t) \) for \( t \in (0,1] \) (see, (16) in [1]).

**Lemma 2.3** One has
\[ \rho(s) \geq \left(1 + \frac{\log(1+2s)}{\log(q)}\right)^{-1}. \]

The following property lemma shows that the proposed parametric kernel function (6) has the exponential convexity, see, e.g., [4,12].

**Lemma 2.4** (Lemma 2.1 in [4]) Let \( t_1 > 0 \) and \( t_2 > 0 \). Then
\[ \phi(\sqrt{t_1 t_2}) \leq \frac{1}{2} (\phi(t_1) + \phi(t_2)). \]

According to the kernel function \( \phi(t) \) given by (6), the barrier function \( \Psi(v) \) is defined by
\[ \Psi(v) := \text{Tr}(\phi(v)), \quad v \in \text{int}V. \]

It follows that
\[ \Psi(v) = \sum_{i=1}^{N} \left( \phi(\lambda_{\min}(v^{(i)})) + \phi(\lambda_{\max}(v^{(i)})) \right) \geq 0. \]

**Lemma 2.5** Let \( x, s, v \in \text{int}V \),
\[ \text{Tr}(v \circ v) = \text{Tr}(x \circ s) \]
and
\[ \text{det}(v \circ v) = \text{det}(x)\text{det}(s), \]
Then
\[ \Psi(v) \leq \frac{\Psi(x) + \Psi(s)}{2}. \]

The following theorem provides an estimate for the effect of a \( \mu \)-update on the value of \( \Psi(V) \), which is a reformulation of Theorem 3.2 in [4].

**Theorem 2.1** Let \( v \in \text{int}V \) and \( \beta \geq 1 \). Then
\[ \Psi(\beta v) \leq 2N \phi\left( \beta \tilde{n}\left( \frac{\Psi(v)}{2N} \right) \right). \]
Corollary 2.1 Let \( 0 \leq \theta < 1 \) and \( v_+ = \frac{v}{\sqrt{1-\theta}} \). If \( \Psi(v) \leq \tau \), then

\[
\Psi(v_+) \leq 2N\varphi\left(\frac{n(\tau/2N)}{\sqrt{1-\theta}}\right).
\]

**Proof** With \( \beta = \frac{1}{\sqrt{1-\theta}} \geq 1 \) and \( \Psi(v) \leq \tau \), the result follows immediately from Theorem 2.1. This completes the proof.

The norm-based proximity measure \( \sigma(v) \) is given by

\[
\sigma(v) := \sqrt{\frac{1}{N} \sum_{i=1}^{N} \|\varphi'(v^{(i)})\|^2}, \quad v \in \text{int} V.
\]

(12)

It follows that

\[
2\sigma(v) = \sqrt{\sum_{i=1}^{N} \left(\varphi'(\lambda_{\min}(v^{(i)}))^2 + \varphi'(\lambda_{\max}(v^{(i)}))^2\right)} \geq 0.
\]

(13)

The lower bound on \( \sigma(v) \) in terms of \( \Psi(v) \) can be obtained from the following theorem, which is a reformulation of Theorem 4.8 in [4].

**Theorem 2.2** Let \( v \in \text{int} V \). Then

\[
\sigma(v) \geq \frac{1}{2} \psi'(\tilde{n}(\Psi(v))).
\]

**Corollary 2.2** Let \( v \in \text{int} V \). Then

\[
\sigma(v) \geq \frac{\sqrt{2\Psi(v)}}{2}.
\]

**Proof** We have

\[
\sigma(v) \geq \frac{1}{2} \psi'(\tilde{n}(\Psi(v))) \geq \frac{1}{2} \left(\sqrt{1+2\Psi(v)} - \frac{1}{\sqrt{1+2\Psi(v)}}\right) \geq \frac{1}{2} \left(\sqrt{1+2\Psi(v)} - 1\right) = \frac{\Psi(v)}{1+\sqrt{1+2\Psi(v)}} \geq \frac{\sqrt{2\Psi(v)}}{2}.
\]

This completes the proof.

3. THE GENERIC KERNEL-BASED IPAS FOR SOCO

3.1 The central path for SOCO

Without loss of generality, we assume that SOCO problems satisfies interior-point condition (IPC), i.e., there exists \((x^0, y^0, s^0) \in \text{int} V \times R^n \times \text{int} V\) such that

\[
Ax^0 = b, \quad A^T y^0 + s^0 = c.
\]

(14)

Then the optimality condition for SOCO is equivalent to the following system.

\[
\begin{cases}
Ax = b, \ x \in V, \\
A^T y + s = c, \ s \in V, \\
x \circ s = 0.
\end{cases}
\]

(15)
The standard approach is to replace the third equation in (15), the so-called complementarity condition for SOCO, by the parameterized equation \( x \circ s = \mu e \), with \( \mu > 0 \) and \( e = (e^{(1)}, \ldots, e^{(n)})^T \) is the unit element in \( (R^n, \circ) \). This yields to the following system

\[
\begin{align*}
Ax &= b, \ x \in V, \\
A^Ty + s &= c, \ s \in V, \\
x \circ s &= \mu e.
\end{align*}
\]

The system (16) has a unique solution \((x(\mu), y(\mu), s(\mu))\) [12].

Let \( x(\mu) \) and \( (y(\mu), s(\mu)) \) be the \( \mu \)-center of the primal problem and its dual problem, respectively. We call the set of \( \mu \)-centers the central path for the IPAs. If \( \mu \to 0 \), then there exists the limit of the central path and since the limit points satisfy the complementarity condition, i.e., \( x \circ s = 0 \), the limit point yields an optimal solution for SOCO.

### 3.2 The NT-search directions for SOCO

From (16), we obtain, by using Newton’s method,

\[
\begin{align*}
A\Delta x &= 0, \\
A^T\Delta y + \Delta s &= 0, \\
x \circ \Delta s + s \circ \Delta x &= \mu e - x \circ s.
\end{align*}
\]

Due to the fact that \( L(x)L(s) \neq L(s)L(x) \) in general. The system (17) doesn’t have a unique solution in general. For overcome this difficulty, we consider the following system

\[
\begin{align*}
Ax &= b, \ x \in V, \\
A^Ty + s &= c, \ s \in V, \\
P(u)x \circ P(u^{-1})s &= \mu e.
\end{align*}
\]

This due to the fact that \( P(u)x \circ P(u^{-1})s = \mu e \Leftrightarrow x \circ s = \mu e \) (see, e.g., Lemma 28 in [14]). Applying Newton’s method again, we have

\[
\begin{align*}
A\Delta x &= 0, \\
A^T\Delta y + \Delta s &= 0, \\
P(u)x \circ P(u^{-1})\Delta s + P(u^{-1})s \circ P(u)\Delta x &= \mu e - P(u)x \circ P(u^{-1})s.
\end{align*}
\]

In this paper, we choose the Nesterov and Todd scaling scheme [12] due to the fact that it is a powerful tool to transform the primal variable \( x \) and the dual variable \( s \) into the same space: the so-called \( v \)-space. Thus it leads to the system (19) into a well-defined one.

Let \( u := w^{-\frac{1}{2}} \). The scaled vector is given by

\[
v := \frac{P(w)^{\frac{1}{2}}x}{\sqrt{\mu}} = \frac{P(w)^{\frac{1}{2}}s}{\sqrt{\mu}}.
\]

Furthermore, we define the scaled search directions as follows

\[
d_x := \frac{P(w)^{\frac{1}{2}}\Delta x}{\sqrt{\mu}}, \ d_s := \frac{P(w)^{\frac{1}{2}}\Delta s}{\sqrt{\mu}}.
\]
\[
\begin{equation}
\begin{bmatrix}
\bar{A}d_s = 0, \\
\bar{A}^T \Delta y + d_s = 0, \\
d_s + d_y = v^{-1} - v,
\end{bmatrix}
\end{equation}
\]

where \( \bar{A} := \frac{1}{\sqrt{\mu}} AP(w)^{\frac{1}{2}} \). The system (22) has a unique scaled search direction [12]. A crucial observation is that the right-hand side \( v^{-1} - v \) in the second equation of the system (22) equals minus the gradient \( \nabla \Psi(v) \) of the classical logarithmic barrier function

\[
\Psi(v) := \text{Tr}(\varphi(v)) = \sum_{i=1}^{N} \left( \varphi_i(\lambda_i(v)) \right),
\]

where

\[
\varphi_i(t) = \frac{t^2 - 1}{2} - \log t.
\]

We can conclude that the system (19) is equivalent to the following system.

\[
\begin{equation}
\begin{bmatrix}
\bar{A}d_s = 0, \\
\bar{A}^T \Delta y + d_s = 0, \\
d_s + d_y = -\nabla \Psi(v).
\end{bmatrix}
\end{equation}
\]

This means that the logarithmic barrier function essentially determines the classical Nesterov and Todd search direction.

For the purpose of obtaining the new search directions, we replace \( -\nabla \Psi(v) \) by \( \nabla \Psi(v) \) with the function \( \varphi(t) \) given by (6). Then

\[
\begin{equation}
\begin{bmatrix}
\bar{A}d_s = 0, \\
\bar{A}^T \Delta y + d_s = 0, \\
d_s + d_y = -\nabla \Psi(v).
\end{bmatrix}
\end{equation}
\]

Since the system (26) has the same coefficient matrix with the system (19), then the system (26) also has a unique solution. The new search directions \( d_s \) and \( d_y \) are obtained by solving (26). It follows from (21) that

\[
\Delta x = \sqrt{\mu P(w)^{\frac{1}{2}}} \Delta x, \quad \Delta y = \sqrt{\mu P(w)^{\frac{1}{2}}} \Delta y, \quad \Delta s = \sqrt{\mu P(w)^{\frac{1}{2}}} \Delta s.
\]

If \( (x, y, s) \neq (x(\mu), y(\mu), s(\mu)) \), then \( (\Delta x, \Delta y, \Delta s) \) is nonzero. By taking a default step size \( \alpha \) along the search directions, we get the new iteration point as follows:

\[
x_+ := x + \alpha \Delta x, \quad y_+ := y + \alpha \Delta y, \quad s_+ := s + \alpha \Delta s.
\]

Furthermore, we can conclude that

\[
\Psi(v) = 0 \iff \sigma(v) \iff v = e \iff xs = \mu e.
\]

This means that \( \Psi(v) \) can be used to estimate the distance between \( (x, y, s) \) and \( (x(\mu), y(\mu), s(\mu)) \).

### 3.3 The kernel-based IPAs for SOCO

Let \( \mu^0 = 1 \) and \( 0 < \tau < 1 \) be a threshold value. At the beginning of the algorithm, we suppose that \( (x^0, y^0, s^0) \in \text{int} V \times R^m \times \text{int} V \) are feasible and \( \Psi(x^0, s^0; \mu^0) \leq \tau \). Let \( 0 < \theta < 1 \). Then, we have a new outer iteration by updating \( \mu \) with \( (1-\theta) \mu \). If \( \Psi(v) \) is large than \( \tau \), then we have a new inner iteration. First, we compute the
scaled search direction \((d_x, d_y, d_s)\) from (26), so that \(\Delta x\) and \(\Delta s\) are calculated from (27). Finally, the new iterates is computed from (28). This process is repeated until we get the new iterates such that \(\Psi(x, y, s) \leq \tau\). Then we update \(\mu\) and use Newton’s method again, and so on, only if \(N\mu < \varepsilon\). At this stage, we have obtained an \(\varepsilon\) - solution of SOCO is obtained.

The generic kernel-based IPAs for SOCO can be summarized in Figure 1.

**Generic Kernel-Based IPAs for SOCO**

**Inputs:**
- a threshold parameter \(\tau \geq 1\);
- an accuracy parameter \(\varepsilon > 0\);
- a fixed barrier update parameter \(\theta, 0 < \theta < 1\);
- a strictly feasible point \((x^0, s^0)\) and \(\mu^0 = 1\) such that \(\Psi(x^0, s^0; \mu^0) \leq \tau\).

**begin**

\[x := x^0; y := y^0; s := s^0; \mu := \mu^0;\]

**while** \(N\mu \geq \varepsilon\) **do**  

**begin**

\[\mu := (1 - \theta)\mu;\]

**while** \(\Psi(x, s; \mu) > \tau\) **do**  

**begin**

- compute the search directions \((\Delta x, \Delta y, \Delta s)\);
- choose a suitable step size \(\alpha\);
- update \((x, s) := (x, y, s) + \alpha(\Delta x, \Delta y, \Delta s)\).

**end**

**end**

**end**

**Figure 1: Algorithm**

### 4. THE ITERATION COMPLEXITY OF ALGORITHMS

In each inner iteration after the default step, we obtain the new iterates

\[x_\tau = \sqrt{\mu P(w)^{1/2}(v + \alpha d_x)}, \quad s_\tau = \sqrt{\mu P(w)^{-1/2}(v + \alpha d_s)},\]

Then, the scaled iterate \((v_\tau, v + \alpha d_x)\) is obtained from (21). It follows from Lemma 2.1 that the new scaled element \(v_\tau\) is given by

\[v_\tau = P(w_{\tau})^{1/2}P(w)^{1/2}(v + \alpha d_x) = P(w_{\tau})^{1/2}P(w)^{-1/2}(v + \alpha d_s),\]

where

\[w_{\tau} = P(x_{\tau})^{1/2}(P(x_{\tau})^{1/2}s_{\tau})^{-1/2}\].

We can conclude that

\[\text{Tr}\left((v_\tau)^2\right) = \text{Tr}\left((v + \alpha d_x)\otimes(v + \alpha d_s)\right),\]

and
\[ \det \left( (v_+)^2 \right) = \det(v + \alpha d_v) \det(v + \alpha d_s). \]

It follows from Lemma 2.5 that
\[ \Psi(v_+) \leq \frac{\Psi(v + \alpha d_v) + \Psi(v + \alpha d_s)}{2}. \]

Now, we want to choose a default step size such that
\[ f(\alpha) = \Psi(v_+) - \Psi(v) \]

is as small as possible. For this purpose, we define
\[ f_1(\alpha) := \frac{\Psi(v + \alpha d_v) + \Psi(v + \alpha d_s)}{2} - \Psi(v), \]

which is convex due to the fact that \( \Psi(v) \) is convex. Hence, we have
\[ f(\alpha) \leq f_1(\alpha). \]

This means that \( f_1(\alpha) \) gives an upper bound for the decrease of the barrier function \( \Psi(v) \). Moreover, we have \( f(0) = f_1(0) = 0 \). It follows from Lemma 2.10 in [12] that
\[ f_1(\alpha) = \frac{1}{2} \text{Tr} \left( \phi'(v + \alpha d_v) \circ d_v + \phi'(v + \alpha d_s) \circ d_s \right), \]

and
\[ f_1^*(\alpha) = \frac{1}{2} \frac{d^2}{d\alpha^2} \text{Tr}(\phi(v + \alpha d_v) + \phi(v + \alpha d_s)) \leq \frac{\omega_2 \| d_v \|^2 + \omega_1 \| d_s \|^2}{2}, \]

where
\[ \omega_1 = \max \left\{ \left| \phi''(\lambda_{\min}(v + \alpha d_v)) \right|, \left| \phi''(\lambda_{\max}(v + \alpha d_v)) \right|, \frac{\left| \phi'(\lambda_{\max}(v + \alpha d_v)) - \phi'(\lambda_{\min}(v + \alpha d_v)) \right|}{2\|v + \alpha d_v\|_2} \right\}, \]

and
\[ \omega_2 = \max \left\{ \left| \phi''(\lambda_{\min}(v + \alpha d_s)) \right|, \left| \phi''(\lambda_{\max}(v + \alpha d_s)) \right|, \frac{\left| \phi'(\lambda_{\max}(v + \alpha d_s)) - \phi'(\lambda_{\min}(v + \alpha d_s)) \right|}{2\|v + \alpha d_s\|_2} \right\}. \]

The following theorem provide an upper bound of \( f_1^*(\alpha) \) in terms of \( \sigma \) and \( \phi''(\cdot) \).

**Theorem 4.1 (Lemma 3.3 in [6])** Let \( \sigma := \sigma(v) \), Then
\[ f_1^*(\alpha) \leq 2\sigma^2 \phi''(\lambda_{\min}(v) - 2\alpha \sigma). \]

The default step size for the algorithm should be chosen such that \( x_+ \) and \( s_+ \) are feasible and \( f(\alpha) \), i.e., \( \Psi(v_+) - \Psi(v) \), decreases sufficiently. For the details we leave it for the interested readers (see, e.g., [5,6]). Following the strategy considered in [4], we briefly recall how to choose the default step size. Suppose that the step size \( \alpha \) satisfies
\[ -\phi'(\lambda_{\min}(v) - 2\alpha \sigma) + \phi'(\lambda_{\min}(v)) \leq 2\sigma. \]

Then \( f_1(\alpha) \leq 0 \). The largest possible value of the step size of \( \alpha \) satisfying (35) is given by
\[ \tilde{\alpha} := \frac{1}{2\sigma} \left( \rho(\sigma) - \rho(2\sigma) \right). \]

This implies that
Furthermore, we have
\[
\bar{\alpha} \geq \frac{1}{1 + (1 + 4\sigma)(2 + \log q) \left(1 + \frac{\log(1 + 4\sigma)}{\log q}\right)^2}.
\]  

In the sequel, we use
\[
\tilde{\alpha} := \frac{1}{1 + (1 + 4\sigma)(2 + \log q) \left(1 + \frac{\log(1 + 4\sigma)}{\log q}\right)^2}
\]  
as the default step size.

**Lemma 4.1** Let \( \alpha \leq \bar{\alpha} \). Then
\[
f(\alpha) \leq -\alpha\sigma^2.
\]

**Proof** Since \( f'_1(\alpha) \) is a twice differentiable convex function with \( f'_1(0) = 0 \), and \( f'_1(0) = -2\sigma^2 < 0 \), we have, by Lemma A.3 in [4],
\[
f(\alpha) \leq f'_1(\alpha) \leq -\alpha\sigma^2.
\]
This completes the proof.

The following theorem shows that the default step size yields sufficient decrease of the barrier function during each inner iteration.

**Theorem 4.2** Let \( \Psi_0 \geq \tau \geq 3 \). Then
\[
f(\bar{\alpha}) \leq \frac{1}{20(2 + \log q) \left(1 + \frac{\log(1 + \sqrt{\Psi_0})}{\log q}\right)^2} \Psi^\frac{1}{2}(v),
\]

**Proof** Let \( \Psi_0 \geq \tau \geq 3 \). Then \( \sigma \geq 1 \) and \( \sqrt{\Psi_0} \leq \sqrt{2\sigma} \leq 2\sigma \). From Lemma 4.1 with (38) and Corollary 2.2, we have
\[
f(\bar{\alpha}) \leq -\bar{\alpha}\sigma^2 \leq -\frac{\sigma^2}{10\sigma(2 + \log q) \left(1 + \frac{\log(1 + 4\sigma)}{\log q}\right)^2} \leq -\frac{\sigma}{10(2 + \log q) \left(1 + \frac{\log(1 + 4\sigma)}{\log q}\right)^2}
\]
\[
\leq \frac{1}{20(2 + \log q) \left(1 + \frac{\log(1 + \sqrt{\Psi_0})}{\log q}\right)^2} \Psi^\frac{1}{2}(v).
\]
This completes the proof.

At the start of an outer iteration and just before updating the parameter \( \mu \), we have \( \Psi(v) \leq \tau \). It follows from Corollary 2.2 that the value of \( \Psi(v) \) exceeds from the threshold \( \tau \) after updating of \( \mu \).

Therefore, we need to count how many inner iterations are required to return to the situation where \( \Psi(v) \leq \tau \). We denote the value of \( \Psi(v) \) after the \( \mu \)-update as \( \Psi_0 \), the subsequent values in the same outer iteration are denoted as \( \Psi_k \) with \( k = 1, \ldots, K \), where \( K \) denotes the total number of inner iterations in the outer iteration.

It follows from (6) that
Furthermore, we have, by Corollary 2.2,
\[
\Psi_0 \leq 2N\phi\left(\frac{p(\tau/2N)}{\sqrt{1-\theta}}\right) \leq 2N\phi\left(\frac{1+\sqrt{\tau/N}}{\sqrt{1-\theta}}\right) \leq \frac{2N\left(\sqrt{2N\tau + \tau/\theta} + N\theta\right)}{1-\theta} = O(N).
\]
(40)
We have, by Theorem 4.2,
\[
\Psi_{k+1} \leq \Psi_k - \beta(\Psi_k)^{1+\gamma}, \quad k = 0, 1, \ldots, K-1,
\]
where
\[
\beta = \frac{1}{20(2+\log q)\left(1 + \frac{\log(1+\sqrt{\Psi_0})}{\log q}\right)}, \quad \gamma = \frac{1}{2}.
\]
From Lemma A.2 in [4] and (41), we have the following lemma, which gives an estimate for the number of inner iterations between two successive barrier parameter updates.

**Lemma 4.2** One has
\[
K \leq 40(2+\log q)\left(1 + \frac{\log(1+\sqrt{\Psi_0})}{\log q}\right)\left(\Psi_0\right)^{\frac{1}{2}}.
\]
It should be noted that an upper bound of the number of outer iterations is bounded above by [13]
\[
\left[\frac{1}{\theta} \log \frac{N}{\varepsilon}\right].
\]
By multiplying the number of outer iterations and the number of inner iterations, we obtain an upper bound for the total number of iterations, namely,
\[
\frac{40(2+\log q)}{\theta}\left(1 + \frac{\log\left(1+\sqrt{2N\tau + \tau/\theta} + N\theta\right) / \left(1-\theta\right)}{\log q}\right)\left(\frac{2\sqrt{N\tau + \tau + N\theta}}{1-\theta}\right)^{\frac{1}{2}} \log \frac{N}{\varepsilon}.
\]
**Theorem 4.3** For large-update methods, namely, \(\theta = \Theta(1), \tau = O(N)\). Then the iteration complexity becomes
\[
O\left(\sqrt{N} \log N \log(N / \varepsilon)\right)
\]
which matches the currently best well-know iteration complexity for large-update methods.

**Proof** It follows from (40) that \(\Psi_0 \leq O(N)\). By choosing \(q = 1 + O(N)\), we can easily obtain the best total iteration bound. This completes the proof.

5. **CONCLUSIONS**

In this paper, we have shown that kernel-based long-step IPAs for LO can be generalized to the SOCO case. The symmetrization of the search directions used in this paper is based on the Nesterov and Todd scaling scheme. The iteration complexity is bounded above by
\[
O\left(\sqrt{N} \log N \log(N / \varepsilon)\right),
\]
which matches the currently best result of iteration bounds for large-update methods.
ACKNOWLEDGMENTS

This work was supported by University Students’ Innovative Training Program of Shanghai (Nos. CS1721001 and CS1621001).

REFERENCES


