J Glob Optim (2007) 39:555–575
DOI 10.1007/s10898-007-9156-y

ORIGINAL PAPER

Entropy-like proximal algorithms based on a second-order homogeneous distance function for quasi-convex programming

Shaohua Pan · **Jein-Shan Chen**

Received: 13 August 2006 / Accepted: 12 March 2007 / Published online: 25 April 2007

Abstract We consider two classes of proximal-like algorithms for minimizing a proper lower semicontinuous quasi-convex function $f(x)$ subject to non-negative constraints $x \geq 0$. The algorithms are based on an entropy-like second-order homogeneous distance function. Under the assumption that the global minimizer set is nonempty and bounded, we prove the full convergence of the sequence generated by the algorithms, and furthermore, obtain two important convergence results through imposing certain conditions on the proximal parameters. One is that the sequence generated will converge to a stationary point if the proximal parameters are bounded and the problem is continuously differentiable, and the other is that the sequence generated will converge to a solution of the problem if the proximal parameters approach to zero. Numerical experiments are done for a class of quasi-convex optimization problems where the function $f(x)$ is a composition of a quadratic convex function from $\mathbb{R}^n$ to $\mathbb{R}$ and a continuously differentiable increasing function from $\mathbb{R}$ to $\mathbb{R}$, and computational results indicate that these algorithms are very promising in finding a global optimal solution to these quasi-convex problems.

Keywords Proximal-like method · Entropy-like distance · Quasi-convex programming

Shaohua Pan work was partially supported by the Doctoral Starting-up Foundation (05300161) of GuangDong Province.
Member of Mathematics Division, National Center for Theoretical Sciences, Taipei Office. Jein-Shan Chen work is partially supported by National Science Council of Taiwan.

S. Pan (✉)
School of Mathematical Sciences, South China University of Technology, Guangzhou 510641, China
e-mail: shhpan@scut.edu.cn

J.-S. Chen
Department of Mathematics, National Taiwan Normal University, Taipei 11677, Taiwan
e-mail: jschen@math.ntnu.edu.tw

1 Introduction

The proximal point algorithm for minimizing a convex function $f(x)$ on $\mathbb{R}^n$ generates a sequence $\{x^k\}_{k\in N} \subseteq \mathbb{R}^n$ by the following iterative scheme:

$$x^{k+1} = \operatorname*{argmin}_{x\in\mathbb{R}^n} \left\{ f(x) + \lambda_k \|x - x^k\|^2 \right\}, \tag{1}$$

where λ_k is a sequence of positive numbers and $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^n$. This method, originally introduced by Martinet [15], is based on the Moreau proximal approximation of f (see [16]). The proximal point algorithm was then further developed and studied by Rockafellar [19,20]. Later, several researchers [4,5,7,12,14,23] proposed and studied nonquadratic proximal point algorithm by replacing the quadratic distance in (1) with a Bregman distance or an entropy-like distance. Among others, the entropy-like distance, also called φ-divergence, is defined by

$$d_\varphi(x, y) = \sum_{i=1}^{n} y_i \varphi(x_i/y_i), \tag{2}$$

where $\varphi\colon \mathbb{R} \to (-\infty, +\infty]$ is a closed proper strictly convex function satisfying certain conditions; see [12,13,23,24]. This class of distance-like functions was first proposed by Teboulle [23] in order to define entropy-like proximal maps. A popular choice of φ is the case that $\varphi(t) = t \ln t - t + 1$, for which the corresponding d_φ is exactly the well-known Kullback–Leibler entropy function from statistics [7,8,10,23] and that is the “entropy" terminology stems from.

The proximal-like algorithm based on φ-divergence, originally designed for minimizing a convex function $f(x)$ subject to non-negative constraints $x \geq 0$, consists of a sequence $\{x^k\}_{k\in N} \subseteq \mathbb{R}^n_{++}$ generated by the iterative scheme as follows:

$$\begin{aligned} &x^0 > 0, \\ &x^{k+1} = \operatorname*{argmin}_{x\geq 0} \{ f(x) + \lambda_k d_\varphi(x, x^k) \}. \end{aligned} \tag{3}$$

This class of proximal-like algorithms were studied extensively for convex programming; see [12,13,23,24] and references therein, and particularly, the one with $\varphi(t) = t \ln t - t + 1$ was recently extended to convex semidefinite programs [6] and convex second-order cone programs in a recent manuscript of J.-S. Chen. In fact, the algorithm (3) associated with $\varphi(t) = -\ln t + t - 1$ was first proposed by Eggermont [8]. It is worth to point out that the fundamental difference between (1) and (3) is that the term $d_\varphi(\cdot, \cdot)$ is used in (3) to force the iterates $\{x^k\}_{k\in N}$ to stay in $\mathbb{R}^n_{++}$ which is the interior of the non-negative orthant, namely the algorithm (3) will automatically generate a positive sequence $\{x^k\}_{k\in N} \subseteq \mathbb{R}^n_{++}$.

In this paper, we will focus on two classes of proximal-like algorithms of the form (3) but with a second-order homogeneous distance-like function d_ϕ given by

$$d_\phi(x, y) = \sum_{i=1}^{n} y_i^2 \phi(x_i/y_i), \tag{4}$$

where the kernel ϕ is defined with two types of special φ and a quadratic function. The definition of ϕ and the properties of d_ϕ are given in Sect. 3. This class of algorithms has been studied for convex minimization (see [1,2,22]). However, we in this paper employ these algorithms to solve the following quasi-convex minimization problem:

$$\begin{aligned} &\min\ f(x) \\ &\text{s.t.}\ \ x \geq 0, \end{aligned} \tag{5}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$ is a proper lower semicontinuous quasi-convex function. Since we do not require the convexity of f, the basic iterative scheme for the algorithms is as follows:

$$\begin{aligned} &x^0 > 0, \\ &x^{k+1} \in \operatorname*{argmin}_{x\geq 0} \{f(x) + \lambda_k d_\phi(x, x^k)\}, \end{aligned} \tag{6}$$

where λ_k is same as before. The purpose of this paper is to establish the full convergence of the sequence $\{x^k\}_{k\in N}$ generated by (6) under some mild assumptions for the quasi-convex problem (5), and verify the effectiveness of the algorithms by numerical experiments.

Note that (5) is a special nonconvex optimization problem, and therefore the global optimization methods [11] developed for the general nonconvex optimization problem can be applied for solving it. Nevertheless, we should point out that the design of these global optimization methods is often far more complex than that of the proximal-like method (6).

The rest of this paper is organized as follows. In Sect. 2, we recall some definitions and basic results that will be used in the later sections. In Sect. 3, we present the definition of the kernel ϕ and investigate the properties of d_ϕ. Based on the entropy-like second-order homogeneous distance function d_ϕ, we in Sect. 4 propose two classes of proximal-like algorithms, and prove the full convergence of the sequence generated. In Sect. 5, numerical experiments were done with a specific d_ϕ for a class of continuously differentiable quasi-convex programming problems.

Unless otherwise stated, in this paper, we use the notation $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ to denote the Euclidean inner product and Euclidean norm in $\mathbb{R}^n$, and $\mathbb{R}^n_+$ to represent the non-negative orthant in $\mathbb{R}^n$ with the interior $\mathbb{R}^n_{++}$. For a given differentiable function $f\colon \mathbb{R}^n \to \mathbb{R}$, $\nabla f(x)$ denotes the gradient of f at x, while $(\nabla f(x))_i$ means the ith partial derivative of f with respect to x. In addition, we use $\nabla_1 d_\phi(x, y)$ to denote the partial derivative of d_ϕ with respect to its first component.

2 Basic concepts

In this section, we recall some definitions and basic results which will be used in the subsequent analysis. We start with the definition of Fejér convergence for a sequence.

Definition 2.1 A sequence $\{y^k\}_{k\in N}$ is Fejér convergent to a nonempty set $U \subseteq \mathbb{R}^n$ with respect to a distance-like function $d(\cdot, \cdot)$, if for every $u \in U$, we have $d(u, y^{k+1}) \leq d(u, y^k)$. When d is the Euclidean distance, $\{y^k\}$ is called Fejér convergent to U.

Given an extended real-valued function $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, denote its domain by

$$\operatorname{dom} f := \{x \in \mathbb{R}^n : \ f(x) < +\infty\}$$

and its epigraph by

$$\operatorname{epi} f := \Big\{(x, \beta) \in \mathbb{R}^n \times \mathbb{R} : \ f(x) \leq \beta\Big\}.$$

Then, f is said to be proper if $\operatorname{dom} f \neq \emptyset$ and $f(x) > -\infty$ for any $x \in \operatorname{dom} f$, and f is a lower semicontinuous function if $\operatorname{epi} f$ is a closed subset of $\mathbb{R}^n \times \mathbb{R}$. We next recall the definition of the Fréchet subdifferential; see [18, Chapter 8] and [21, Chapter10].

Definition 2.2 Let $f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function. For each $x \in \operatorname{dom} f$, the Fréchet subdifferential of f at x, denoted by $\hat{\partial} f(x)$, is the set of vectors $s \in \mathbb{R}^n$ such that

$$\liminf_{y \neq x, y \to x} \frac{1}{\|y - x\|} \Big[f(y) - f(x) - \langle s, y - x \rangle \Big] \geq 0. \tag{7}$$

If $x \notin \operatorname{dom} f$, then $\hat{\partial} f(x) = \emptyset$.

The vector s satisfying the inequality (7) is also termed as a *regular subgradient* of f at x (see [21, p. 301]). It is not difficult to see that the inequality (7) is equivalent to

$$f(y) \geq f(x) + \langle s, y - x \rangle + o(\|y - x\|),$$

where

$$\lim_{y \to x} o(\|y - x\|)/\|y - x\| = 0.$$

For the subdifferential $\hat{\partial} f(x)$, the following results hold by direct verifications.

Lemma 2.3 [21, Chapter 8] *Let $f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function and $\hat{\partial} f(x)$ be the subdifferential of f at x. Then,*

(a) *$\hat{\partial} f(x)$ is a closed and convex set.*
(b) *If f is differentiable at x or in a neighborhood of x, then $\hat{\partial} f(x) = \{\nabla f(x)\}$, where $\nabla f(x)$ is the gradient of f.*
(c) *If $g = f + h$ with f finite at x and h differentiable on a neighborhood of x, then $\hat{\partial} g(x) = \hat{\partial} f(x) + \nabla h(x)$.*
(d) *If f has a local minimum at $\bar{x}$, then $0 \in \hat{\partial} f(\bar{x})$.*

To work with differentiable minimization problems, we also need the following definition.

Definition 2.4 Suppose that $f: \mathbb{R}^n \to \mathbb{R}$ is a differentiable function. Then,

(a) For an unconstrained optimization problem of minimizing $f(x)$ over $x \in \mathbb{R}^n$, x^* is called a stationary point if $\nabla f(x^*) = 0$.
(b) For a constrained optimization problem of minimizing $f(x)$ over $x \in C$ where C is nonempty and convex subset of $\mathbb{R}^n$, x^* is called a stationary point if

$$\nabla f(x^*)^T (x - x^*) \geq 0 \quad \text{for all } x \in C.$$

To close this section, we recall the concept of quasi-convexity, strict quasi-convexity and strong quasi-convexity, and briefly discuss general properties of the minimization problem involving the objective function with such properties.

Definition 2.5 Let $f: \mathbb{R}^n \to \mathbb{R}$ be a proper function. Then, f is called quasi-convex if for all $x, y \in \operatorname{dom} f$ and $\beta \in (0, 1)$, there always holds

$$f(\beta x + (1 - \beta) y) \leq \max\{f(x), f(y)\}.$$

It can be proved that any convex function is also quasi-convex, but the converse is not true. For a quasi-convex function, we have the following important property.

Proposition 2.6 *The proper function $f: \mathbb{R}^n \to \mathbb{R}$ is quasi-convex if and only if the level sets $L_f(\alpha) := \{x \in \operatorname{dom} f \mid f(x) \leq \alpha\}$ are convex for every $\alpha \in \mathbb{R}$.*

Definition 2.7 Let f: $\mathbb{R}^n \to \mathbb{R}$ be a proper function. Then, f is called strictly quasi-convex if for all $x, y \in \text{dom} f$ with $f(x) \neq f(y)$, there always holds

$$f(\beta x + (1-\beta)y) < \max\{f(x), f(y)\} \quad \text{for } \forall \beta \in (0, 1).$$

By [3, Lemma 3.5.7], if f is lower semicontinuous and strictly quasi-convex, then f is quasi-convex. For a strictly quasi-convex function, we have the following important result, which implies that every local optimal solution of (5) is also a global optimal solution.

Proposition 2.8 [3, Theorem 3.5.6] *Let $f: \mathbb{R}^n \to \mathbb{R}$ be a proper strictly quasi-convex function. Consider the problem to minimize $f(x)$ subject to $x \in C$, where C is a nonempty convex set in $\mathbb{R}^n$. If $\bar{x}$ is a local optimal solution, then $\bar{x}$ is also a global optimal solution.*

Definition 2.9 Let f: $\mathbb{R}^n \to \mathbb{R}$ be a proper function. Then, f is called strongly quasi-convex if for all $x, y \in \text{dom} f$ with $x \neq y$, there always holds

$$f(\beta x + (1-\beta)y) < \max\{f(x), f(y)\} \quad \text{for } \forall \beta \in (0, 1).$$

It can be shown that every strongly quasi-convex function is strictly quasi-convex, and every strongly quasi-convex function is quasi-convex even without semicontinuity assumption. When $f(x)$ is strongly quasi-convex, the problem (5) has the unique global optimal solution.

3 Distance-like function d_ϕ and its properties

In this section, we present the definition of the kernel ϕ and investigate the properties of the bivariate function d_ϕ induced by ϕ via formula (4). We start with the assumptions on the function φ, needed to define the kernel ϕ. Let φ: $\mathbb{R} \to (-\infty, +\infty]$ be a closed proper convex function with $\text{dom}\varphi \neq \emptyset$ and $\text{dom}\varphi \subseteq [0, +\infty)$. We assume that

(i) φ is twice continuously differentiable on $\text{int}(\text{dom}\varphi) = (0, +\infty)$;
(ii) φ is strictly convex on its domain;
(iii) $\lim_{t \to 0^+} \varphi'(t) = -\infty$;
(iv) $\varphi(1) = \varphi'(1) = 0$ and $\varphi''(1) > 0$.

In the rest of this paper, we denote by Φ the class of functions satisfying (1)–(4).

Given $\varphi \in \Phi$, we define the following two subclasses of Φ:

$$\Phi_1 = \left\{\varphi \in \Phi : \ \varphi''(1)(1 - 1/t) \leq \varphi'(t) \leq \varphi''(1) \ln t, \ \ \forall t > 0\right\} \tag{8}$$

and

$$\Phi_2 = \left\{\varphi \in \Phi : \ \varphi''(1)(1 - 1/t) \leq \varphi'(t) \leq \varphi''(1)(t-1), \ \ \forall t > 0\right\}. \tag{9}$$

Since $\ln t \leq t - 1$ for any $t > 0$ and $\varphi''(1) > 0$, clearly, $\Phi_1 \subseteq \Phi_2 \subseteq \Phi$. The assumptions on Φ_1 and Φ_2 are very mild. It is not hard to verify that the following functions

$$\begin{aligned} \varphi_1(t) &= t \ln t - t + 1, \quad && \text{dom } \varphi = [0, +\infty), \\ \varphi_2(t) &= -\ln t + t - 1, \quad && \text{dom } \varphi = (0, +\infty), \\ \varphi_3(t) &= (\sqrt{t} - 1)^2, \quad && \text{dom } \varphi = [0, +\infty) \end{aligned}$$

are all in Φ_1, and consequently belong to Φ_2. The first example φ_1 plays an important role in the convergence analysis of our first class of algorithms that will be studied in the next section. As mentioned in the introduction, the φ-divergence for $\varphi = \varphi_1$ is exactly the Kullback–Leibler entropy function, given by

$$H(x, y) := d_\varphi(x, y) = \sum_{j=1}^{n} x_j \ln(x_j/y_j) + y_j - x_j, \tag{10}$$

whose domain can be continuously extended to $\mathbb{R}^n_+ \times \mathbb{R}^n_{++}$ by using the convention that $0 \ln 0 = 0$. The following lemma states some useful properties of $H(x, y)$, and since their proofs are elementary by use of (10), we here omit them.

Lemma 3.1 *Let $H(\cdot, \cdot)$ be defined as in (10). Then, we have the following results.*

(a) *The level sets of $H(x, \cdot)$ are bounded for all $x \in \mathbb{R}^n_+$.*
(b) *If $\{y^k\} \subset \mathbb{R}^n_{++}$ converges to $y \in \mathbb{R}^n_+$, then $\lim_{k\to+\infty} H(y, y^k) = 0$.*
(c) *If $\{z^k\} \subset \mathbb{R}^n_+$, $\{y^k\} \subset \mathbb{R}^n_{++}$ are sequences such that $\{z^k\}$ is bounded, $\lim_{k\to+\infty} y^k = y$ and $\lim_{k\to+\infty} H(z^k, y^k) = 0$, then $\lim_{k\to+\infty} z^k = y$.*

With the above assumptions on φ, we now give the definition of the kernel ϕ involved in the function d_ϕ. Given $\varphi \in \Phi$ and the parameters $\mu > 0$ and $\nu \geq 0$, let $\phi : \mathbb{R} \to (-\infty, +\infty]$ be a closed proper convex function defined by

$$\phi(t) := \mu\varphi(t) + \frac{\nu}{2}(t-1)^2. \tag{11}$$

It is not difficult to verify that ϕ satisfies the properties listed in (i)–(iv), and consequently $\phi \in \Phi$. Particularly, ϕ will be strongly convex on its domain if $\nu > 0$. This implies that the objective function of the subproblem (6), i.e., $f(x) + \lambda \sum_{i=1}^{n} (x_i^k)^2 \phi(x_i/x_i^k)$ will be strictly convex on $\mathbb{R}^n_{++}$ if the parameter λ is set to be sufficiently large, although $f(x)$ itself is quasi-convex. That is to say, the proximal term $d_\phi(\cdot, \cdot)$ plays a convexification role in the quasi-convex subproblem (6), and moreover, the convexification role becomes stronger as the parameter λ increases. In fact, from the computational results in Sect. 5, we may see that the proximal term shows a good convexification role for the quasi-convex function $f(x)$, even for a very small λ.

In what follows, we will concentrate on the properties of the bivariate function d_ϕ.

Lemma 3.2 *Given a $\varphi \in \Phi$ and the parameters $\mu > 0, \nu \geq 0$, and let ϕ be the kernel defined by (11) and $d_\phi(\cdot, \cdot)$ be the function induced by ϕ via formula (4). Then,*

(a) *d_ϕ is a homogeneous function of order 2, i.e., $d_\phi(\alpha x, \alpha y) = \alpha^2 d_\phi(x, y)$ for $\forall \alpha > 0$.*
(b) *For a fixed $y \in \mathbb{R}^n_{++}$, the function $d_\phi(\cdot, y)$ is strictly convex over $\mathbb{R}^n_{++}$. If, in addition, $\nu > 0$, then $d_\phi(\cdot, y)$ is strongly convex on $\mathbb{R}^n_{++}$.*
(c) *For any $(x, y) \in \mathbb{R}^n_{++} \times \mathbb{R}^n_{++}$, $d_\phi(x, y) \geq 0$, and $d_\phi(x, y) = 0$ if and only if $x = y$.*
(d) *For any fixed $z \in \mathbb{R}^n_{++}$, the level sets $L(z, \gamma) := \{x \in \mathbb{R}^n_{++} : d_\phi(x, z) \leq \gamma\}$ are bounded for all $\gamma \geq 0$.*
(e) *If $\varphi \in \Phi_1$ or Φ_2, and $\{y^k\}_{k\in N} \subseteq \mathbb{R}^n_{++}$ converges to $\bar{y} \in \mathbb{R}^n_+$, then for any fixed $x \in \mathbb{R}^n_{++}$, the sequence $\{d_\phi(x, y^k)\}_{k\in N}$ is bounded.*

Proof The properties in (a) and (b) are clear from the definition of d_ϕ given by (4).
(c) Note that $\phi(t)$ is strictly convex and moreover $\phi'(1) = \mu\varphi'(1) = 0$ due to (iv). Hence,

$$\phi(t) \geq \phi(1) = 0 \quad \text{and} \quad \phi(t) = 0 \quad \text{iff } t = 1.$$

This implies that $d_\phi(x, y) \geq 0$ for $\forall(x, y) \in \mathbb{R}^n_{++} \times \mathbb{R}^n_{++}$, and $d_\phi(x, y) = 0$ iff $x = y$.

(d) To prove the result, it is enough to consider the one-dimensional case, i.e., to show that $h_\zeta(t) := \zeta^2\phi(t/\zeta)$ for $\zeta > 0$ has bounded level sets, which in turn is equivalent to showing that ϕ has bounded level sets. Note that $\{t : \phi(t) \le 0\} = \{1\}$. Therefore, the conclusion follows from [18, Corrollary 8.7.1].
(e) From the definitions of ϕ and d_ϕ, we have that

$$d_\phi(x, y^k) = \sum_{i=1}^{n}\left[\mu(y_i^k)^2\varphi\left(\frac{x_i}{y_i^k}\right) + \frac{\nu}{2}(y_i^k)^2\left(\frac{x_i}{y_i^k} - 1\right)^2\right]$$
$$= \sum_{i=1}^{n}\left[\mu(y_i^k)^2\varphi(x_i/y_i^k) + \frac{\nu}{2}(x_i - y_i^k)^2\right].$$

If $\varphi(t)$ is bounded above for any $t > 0$, then the conclusion is obvious. Otherwise, we discuss the following two cases:
Case (1) $\bar{y}_i > 0$ for each $i \in \{1, 2, \dots, n\}$. Since $\{y_i^k\}_{k\in N} \to \bar{y}_i$ for each i, the proof follows directly from the continuity of φ.
Case (2) there exists an index $i_0 \in \{1, 2, \dots, n\}$ such that $\bar{y}_{i_0} = 0$. By the given assumptions and Case (1), it suffices to prove that the sequence $\{(y_{i_0}^k)^2\varphi(x_i/y_{i_0}^k)\}$ is bounded above. For any $k \in N$, using the convexity of φ and the fact that $\varphi(1) = 0$, we have that

$$0 \ge \varphi(x_i/y_{i_0}^k) + \varphi'(x_i/y_{i_0}^k)\left(1 - x_i/y_{i_0}^k\right).$$

Multiplying the inequality with $(y_{i_0}^k)^2$ readily yields that

$$(y_{i_0}^k)^2\varphi(x_i/y_{i_0}^k) \le (y_{i_0}^k)^2\varphi'(x_i/y_{i_0}^k)\left(x_i/y_{i_0}^k - 1\right) = (y_{i_0}^k)\varphi'(x_i/y_{i_0}^k)\left(x_i - y_{i_0}^k\right),$$

which in turn implies that

$$(y_{i_0}^k)^2\varphi(x_i/y_{i_0}^k) \le \left|(y_{i_0}^k)\varphi'(x_i/y_{i_0}^k)(x_i - y_{i_0}^k)\right|.$$

If $\varphi \in \Phi_2$, then it follows from (9) that

$$\varphi''(1)y_{i_0}^k(1 - y_{i_0}^k/x_i) \le y_{i_0}^k\varphi'(x_i/y_{i_0}^k) \le \varphi''(1)(x_i - y_{i_0}^k).$$

Combining the last two inequalities immediately gives that

$$(y_{i_0}^k)^2\varphi(x_i/y_{i_0}^k) \le \max\left\{\varphi''(1)\left(x_i - y_{i_0}^k\right)^2, \varphi''(1)\frac{y_{i_0}^k}{x_i}\left(x_i - y_{i_0}^k\right)^2\right\}.$$

This together with the given assumptions shows that $\{(y_{i_0}^k)^2\varphi(x_i/y_{i_0}^k)\}$ is bounded above for any $\varphi \in \Phi_2$, and consequently the sequence $\{d_\phi(x, y^k)\}_{k\in N}$ is bounded. Noting that $\Phi_1 \subseteq \Phi_2$, the sequence $\{d_\phi(x, y^k)\}_{k\in N}$ is also bounded for $\varphi \in \Phi_1$. □

Lemma 3.2 (a)–(c) state that d_ϕ defined by (4) is a convex second-order homogeneous distance-like function. Thus, in analogy with the Euclidean distance, we can define the projection of a point y, denoted by $\hat{x}(y)$, to a closed convex set $S \subseteq \mathbb{R}^n$ with respect to d_ϕ, which is characterized as the solution of the following problem

$$\inf\left\{d_\phi(x, y) : x \in S\right\}. \tag{12}$$

The existence of $\hat{x}(y)$ is guaranteed by Lemma 3.2 (d). For this projection, we have the following similar results to the Euclidean projection.

Lemma 3.3 *Let S be a closed convex subset of $\mathbb{R}^n$ and $y \in \mathbb{R}^n$ be a point not in S. Then $\hat{x}(y)$ is the projection of y on S with respect to d_ϕ if and only if*

$$\left\langle x - \hat{x}(y),\ -\nabla_1 d_\phi(\hat{x}(y), y)\right\rangle \le 0,\ \ \forall x \in S. \tag{13}$$

Proof Note that problem (12) is equivalent to $\inf\{d_\phi(x, y) + \delta(x \mid S) : x \in \mathbb{R}^n\}$, where $\delta(\cdot \mid S)$ denotes the indicator function of the set S. By [19, Theorem 27.4], $\hat{x}(y)$ solves the unconstrained optimization problem if and only if the inequality (13) holds. Thus, the proof is completed. □

Finally, we present a favorable property of d_ϕ with $\varphi \in \Phi_1$ or Φ_2, which will play a crucial role in the convergence analysis of algorithms to be studied in the next section.

Lemma 3.4 *Given a $\varphi \in \Phi$ and the parameters $\mu > 0, \nu \ge 0$, and let ϕ be the kernel defined as in (11). Then, for any $a, b \in \mathbb{R}^n_{++}$ and $c \in \mathbb{R}^n_{+}$, we have the following results:*

(a) *If $\nu = 0$ and $\varphi \in \Phi_1$, then $\langle c - b, \nabla_1 d_\phi(b, a)\rangle \le \mu\varphi''(1)\max_{1\le j\le n}\{a_j\}[H(c, a) - H(c, b)]$.*
(b) *If $\nu \ge \mu\varphi''(1) > 0$ and $\varphi \in \Phi_2$, then $\langle c - b, \nabla_1 d_\phi(b, a)\rangle \le \theta(\|c - a\|^2 - \|c - b\|^2)$ with $\theta = (\nu + \mu\varphi''(1))/2$.*

Proof (a) Since $\varphi \in \Phi_1$, we have from (8) that

$$\varphi'(t) \le \varphi''(1)\ln t \quad \text{for any } t > 0.$$

Setting $t = b_j/a_j$ in the inequality, we then obtain that

$$c_j a_j \varphi'(b_j/a_j) \le c_j a_j \varphi''(1)\ln(b_j/a_j), \quad j = 1, 2, \ldots, n. \tag{14}$$

On the other hand, it follows from (8) that

$$-\varphi'(t) \le -\varphi''(1)(1 - 1/t), \quad \forall t > 0.$$

Substituting $t = b_j/a_j$ into the inequality and multiplying with a_j gives

$$-b_j a_j \varphi'(b_j/a_j) \le a_j \varphi''(1)(a_j - b_j), \quad j = 1, 2, \ldots, n. \tag{15}$$

Define

$$\Psi'(a, b) := (a_1\varphi'(b_1/a_1), \ldots, a_n\varphi'(b_n/a_n))^T, \quad \forall a, b \in \mathbb{R}^n_{++}.$$

Then, adding the inequalities (14) and (15) and summing over $j = 1, \ldots, n$ gives

$$\begin{aligned}\left\langle c - b, \Psi'(a, b)\right\rangle &\le \varphi''(1)\left[\sum_{j=1}^n a_j\left(c_j\ln(b_j/a_j) + a_j - b_j\right)\right]\\ &\le \varphi''(1)\max_{1\le j\le n}\{a_j\}\left[\sum_{j=1}^n c_j\ln(b_j/a_j) + a_j - b_j\right]\\ &= \varphi''(1)\max_{1\le j\le n}\{a_j\}\left[H(c, a) - H(c, b)\right].\end{aligned}$$

Note that $\nabla_1 d_\phi(b, a) = \mu\Psi'(a, b)$, and hence we obtain the result from the last inequality.
(b) The proof is similar to [2, Lemma 3.4]. For completeness, we here include it. Since $\varphi \in \Phi_2$, the inequality (15) still holds. On the other hand, we have from (9) that

$$\varphi'(t) \le \varphi''(1)(t - 1), \quad \forall t > 0.$$

Substituting $t = b_j/a_j$ into the above inequality leads to

$$c_j a_j \varphi'(b_j/a_j) \le c_j a_j \varphi''(1)(b_j/a_j - 1) = \varphi''(1)c_j(b_j - a_j), \quad j = 1, 2, \ldots, n. \quad (16)$$

Adding the two inequalities (15) and (16), summing over $j = 1, 2, \ldots, n$, and using the definition of $\Psi'(a, b)$, we obtain

$$\langle c - b, \Psi'(a, b)\rangle \le \varphi''(1) \sum_{j=1}^{n} \Big[c_j(b_j - a_j) + a_j(a_j - b_j)\Big] = \varphi''(1)\langle c - a, b - a\rangle.$$

Note that $\nabla_1 d_\phi(b, a) = \mu\Psi'(a, b) + \nu(b - a)$. Then, the last inequality implies that

$$\langle c - b, \nabla_1 d_\phi(a, b)\rangle \le \mu\varphi''(1)\langle c - a, b - a\rangle + \nu\langle c - b, b - a\rangle. \quad (17)$$

Using the identities

$$\langle c - a, b - a\rangle = (1/2)(\|c - a\|^2 - \|c - b\|^2 + \|b - a\|^2)$$

and

$$\langle c - b, b - a\rangle = (1/2)(\|c - a\|^2 - \|c - b\|^2 - \|b - a\|^2)$$

we then from (17) obtain

$$\begin{aligned}\langle c - b, \nabla_1 d_\phi(b, a)\rangle &\le \theta(\|c - a\|^2 - \|c - b\|^2) - \frac{1}{2}(\nu - \mu\varphi''(1))\|b - a\|^2 \\ &\le \theta(\|c - a\|^2 - \|c - b\|^2),\end{aligned}$$

where the second inequality is due to $\nu \ge \mu\varphi''(1)$. Thus, the proof is completed. □

4 Interior proximal-like methods

In this section, we consider two classes of proximal-like algorithms based on the second-order homogeneous function d_ϕ for the quasi-convex optimization problem (5). The two kinds of algorithms are described as follows, where the RIPM was first proposed by Auslender et al. [2] for convex minimization problems subject to non-negative constraints.

Interior Proximal Method (IPM) *Let ϕ be defined as in* (11) *with $\mu > 0, \nu = 0$ and $\varphi \in \Phi_1$. Generate the sequence $\{x^k\}_{k\in N}$ by the iterative scheme (6).*

Regularized Interior Proximal Method (RIPM) *Let ϕ be defined as in* (11) *with $\nu \ge \mu\varphi''(1) > 0$ and $\varphi \in \Phi_2$. Generate the sequence $\{x^k\}_{k\in N}$ by the iterative scheme* (6).

To establish the convergence of IPM and RIPM, throughout this section, we make the following assumptions for the quasi-convex optimization problem (5):

(A1) $\operatorname{dom} f \cap \mathbb{R}^n_{++} \neq \emptyset$.
(A2) The optimal set of problem (5), denoted by $\mathcal{X}^*$, is nonempty and bounded.

In what follows, we concentrate on the convergence of IPM and RIPM. We first prove that they are well-defined, which is a direct consequence of the following lemma.

Lemma 4.1 *Given $\mu > 0, \nu \ge 0$ and $\varphi \in \Phi$, and let ϕ and d_ϕ be defined as in* (11) *and* (4)*, respectively. Then, under assumptions* (A1) *and* (A2)*, the sequence $\{x^k\}_{k\in N}$ generated by the iterative scheme* (6) *is well defined.*

Proof The proof proceeds by induction. Clearly, when $k = 0$, the conclusion holds due to (6). Assume that x^k is well defined. Let f^* be the optimal value of problem (5), then

$$f(x) + \lambda_k d_\phi(x, x^k) \geq f^* + \lambda_k d_\phi(x, x^k) \quad \text{for all } x \in \mathbb{R}^n_{++}. \tag{18}$$

Let $F_k(x) := f(x) + \lambda_k d_\phi(x, x^k)$ and denote its level sets by

$$L_{F_k}(\gamma) := \{x \in \mathbb{R}^n_{++} : F_k(x) \leq \gamma\} \quad \text{for all } \gamma \in \mathbb{R}.$$

Then, the inequality in (18) implies that $L_{F_k}(\gamma) \subseteq L(x^k, \lambda_k^{-1}(\gamma - f^*))$. By Lemma 3.2 (c), the level sets $L(x^k, \lambda_k^{-1}(\gamma - f^*))$ are bounded for any $\gamma \geq f^*$, and consequently, the sets $L_{F_k}(\gamma)$ are bounded for any $\gamma \geq f^*$. Whereas for any $\gamma \leq f^*$, we have $L_{F_k}(\gamma) \subseteq \mathcal{X}^*$, which are obviously bounded due to assumption (A2). The two sides show that the level sets of the function $F_k(x)$ are bounded. Also, $F_k(x)$ is lower semicontinuous on $\operatorname{dom} f$. Hence, the level sets of $F_k(x)$ are compact. Now, using the lower semicontinuity of $F_k(x)$ and the compactness of its level sets, we have that $F_k(x)$ has a global minimum which may not be unique due to the nonconvexity of f. In such case, x^{k+1} can be arbitrarily chosen among the set of minimizers of $F_k(x)$. □

Next, we investigate the properties of the sequence $\{x^k\}_{k\in N}$ generated by IPM and RIPM. To this end, we define the following set

$$U := \left\{x \in \mathbb{R}^n_+ \mid f(x) \leq \inf_{k\in N} f(x^k)\right\}.$$

From assumptions (A1)–(A2) and Proposition 2.6, U is a nonempty closed convex set.

Lemma 4.2 *Let $\{\lambda_k\}_{k\in N}$ be an arbitrary sequence of positive numbers and $\{x^k\}_{k\in N}$ be the sequence generated by IPM. Then, under assumptions* (A1)–(A2)*,*

(a) *$\{f(x^k)\}_{k\in N}$ is a decreasing and convergent sequence.*
(b) *$\{x^k\}_{k\in N}$ is Fejér convergent to the set U with respect to H.*
(c) *For all $x \in U$, the sequence $\{H(x, x^k)\}_{k\in N}$ is convergent.*

Proof (a) From Eq. 6, x^{k+1} is a global optimal solution of the following problem:

$$\min_{x\geq 0} \left\{f(x) + \lambda_k d_\phi(x, x^k)\right\}$$

and consequently, for any $x \in \mathbb{R}^n_+$, it follows that

$$f(x^{k+1}) + \lambda_k d_\phi(x^{k+1}, x^k) \leq f(x) + \lambda_k d_\phi(x, x^k). \tag{19}$$

Setting $x = x^k$ in (19), we then obtain that

$$f(x^{k+1}) + \lambda_k d_\phi(x^{k+1}, x^k) \leq f(x^k) + \lambda_k d_\phi(x^k, x^k) = f(x^k),$$

which means that

$$0 \leq \lambda_k d_\phi(x^{k+1}, x^k) \leq f(x^k) - f(x^{k+1}).$$

Hence, $\{f(x^k)\}_{k\in N}$ is decreasing, and furthermore, convergent due to assumption (A2).
(b) From inequality (19), it follows that for any $x \in U$,

$$d_\phi(x^{k+1}, x^k) \leq d_\phi(x, x^k).$$

This implies that x^{k+1} is the unique projection of x^k on U with respect to d_ϕ. Therefore, by Lemma 3.3, we have that

$$\left\langle x - x^{k+1},\ -\nabla_1 d_\phi(x^{k+1}, x^k)\right\rangle \leq 0, \quad \forall x \in U. \tag{20}$$

On the other hand, applying Lemma 3.4 (a) at the points $c = x$, $a = x^k$, and $b = x^{k+1}$, we then obtain that

$$H(x, x^k) - H(x, x^{k+1}) \geq \frac{\langle x - x^{k+1},\ \nabla_1 d_\phi(x^{k+1}, x^k)\rangle}{\mu\varphi''(1)\max_{1\leq j\leq n}\{x_j^k\}}. \tag{21}$$

Since $\mu\varphi''(1)\max_{1\leq j\leq n}\{x_j^k\} > 0$, using the inequalities (20) and (21) yields that

$$H(x, x^k) \geq H(x, x^{k+1}), \quad \forall x \in U.$$

From Definition 2.1, it follows that $\{x^k\}_{k\in N}$ is Fejér convergent to U with respect to H.
(c) The proof follows from part (b) and the non-negativity of H. □

Lemma 4.3 *Let $\{\lambda_k\}_{k\in N}$ be an arbitrary sequence of positive numbers and $\{x^k\}_{k\in N}$ be the sequence generated by RIPM. Then, under assumptions* (A1) and (A2)*,*

(a) *$\{f(x^k)\}_{k\in N}$ is a decreasing and convergent sequence.*
(b) *$\{x^k\}_{k\in N}$ is Fejér convergent to the set U.*
(c) *For all $x \in U$, the sequence $\{\|x - x^k\|\}_{k\in N}$ is convergent.*

Proof
(a) The proof is similar to that of Lemma 4.2 (a), and we here omit it.
(b) By a similar argument to Lemma 4.2 (b), we can obtain the inequality (20). On the other hand, applying Lemma 3.4 (b) at the points $c = x$, $a = x^k$, and $b = x^{k+1}$ gives

$$\langle x - x^k,\ \nabla_1 d_\phi(x^k, x^{k+1})\rangle \leq \theta(\|x - x^k\|^2 - \|x - x^{k+1}\|^2), \tag{22}$$

where $\theta = (\nu + \mu\varphi''(1))/2$. Since $\theta > 0$, using the inequalities (20) and (22) yields that

$$\|x - x^{k+1}\|^2 \leq \|x - x^k\|^2, \quad \forall x \in U. \tag{23}$$

By Definition 2.1, we thus prove that $\{x^k\}_{k\in N}$ is Fejér convergent to the set U.
(c) The proof follows from part (b) and the non-negativity of $\|x - x^k\|$. □

To now, we have proved that the sequence $\{x^k\}_{k\in N}$ generated by IPM or RIPM is well-defined and satisfies some favorable properties. With these properties, we next establish the convergence results of the proposed algorithms.

Proposition 4.4 *Suppose that assumptions* (A1) and (A2) *are satisfied. Let $\{\lambda_k\}_{k\in N}$ be an arbitrary sequence of positive numbers and $\{x^k\}_{k\in N}$ be the sequence generated by IPM. Then, the sequence $\{x^k\}_{k\in N}$ converges, and furthermore,*

(a) *if there exist $\underline{\lambda}$ and $\bar{\lambda}$ such that $0 < \underline{\lambda} < \lambda_k \leq \bar{\lambda}$ for any k, then*

$$\liminf_{k\to+\infty} g_i^k \geq 0, \quad \lim_{k\to+\infty} g_i^k x_i^k = 0, \quad \forall i = 1, 2, \ldots, n, \tag{24}$$

where $g^k \in \hat{\partial} f(x^k)$ and g_i^k is the ith component of g^k.
(b) *If $\lim_{k\to+\infty}\lambda_k = 0$, then $\{x^k\}_{k\in N}$ converges to a solution of the problem* (5).

Proof We first prove that the sequence $\{x^k\}_{k\in N}$ converges. By Lemma 4.2 (b), $\{x^k\}_{k\in N}$ is Fejér convergent to the set U with respect to H, which implies that

$$\{x^k\}_{k\in N} \subseteq \left\{y \in \mathbb{R}^n_{++} \mid H(x,y) \le H(x,x^0)\right\} \quad \text{for } \forall x \in U.$$

As a consequence, $\{x^k\}_{k\in N}$ is bounded by Lemma 3.1 (a). Thus, there exist an $\bar{x}$ and a subsequence $\{x^{k_j}\}$ of $\{x^k\}_{k\in N}$ converging to $\bar{x}$. From the lower semicontinuity of f,

$$\lim_{j\to+\infty} f(x^{k_j}) \ge f(\bar{x}),$$

which, together with Lemma 4.2 (a), implies that

$$f(\bar{x}) \le f(x^k), \quad \forall k \in N.$$

This shows that $\bar{x} \in U$. By Lemma 4.2 (c), the sequence $\{H(\bar{x},x^k)\}_{k\in N}$ is then convergent. In addition, from Lemma 3.1 (b), we have $\lim_{k\to+\infty} H(\bar{x},x^{k_j}) = 0$. From all the above, we conclude that $\{H(\bar{x},x^k)\}_{k\in N}$ is a convergent sequence with a subsequence converging to 0, and consequently it must converge to 0 itself, i.e., $\lim_{k\to+\infty} H(\bar{x},x^k) = 0$. Using Lemma 3.1 (c) with $z^k = x^k$ and $y^k = \bar{x}$, we thus prove that $\{x^k\}_{k\in N}$ converges to $\bar{x}$.
(a) From the iterative formula (6) and Lemma 2.3 (d), we have that

$$0 \in \hat{\partial}\left(f(x) + \lambda_k d_\phi(x,x^k)\right)(x^{k+1}).$$

Therefore, by Lemma 2.3 (c), there exists $g^{k+1} \in \hat{\partial} f(x^{k+1})$ such that

$$\lambda_k \nabla_1 d_\phi(x^{k+1},x^k) = -g^{k+1},$$

i.e.,

$$\mu\lambda_k x_i^k \varphi'\left(\frac{x_i^{k+1}}{x_i^k}\right) = -g_i^{k+1}, \quad i = 1,2,\ldots,n. \tag{25}$$

Define the index sets

$$I(\bar{x}) := \left\{i \in \{1,2,\ldots,n\} \mid \bar{x}_i > 0\right\} \text{ and } J(\bar{x}) := \left\{i \in \{1,2,\ldots,n\} \mid \bar{x}_i = 0\right\}.$$

We next argue the conclusion by the two cases $i \in I(\bar{x})$ and $i \in J(\bar{x})$.
Case (1) $i \in I(\bar{x})$. In this case, $\lim_{k\to+\infty} x_i^{k+1}/x_i^k = 1$ since $\{x^k\}_{k\in N}$ converges to $\bar{x}$. Using the continuity of φ' and $\varphi'(1) = 0$ and recalling that $0 < \underline{\lambda} \le \lambda_k \le \bar{\lambda}$ for all k, we then obtain from (25) that

$$\lim_{k\to+\infty} g_i^{k+1} = 0, \quad \forall i \in I(\bar{x}). \tag{26}$$

Case (2) $i \in J(\bar{x})$. For every $i \in J(\bar{x})$, we define the following two index sets:

$$J_+^i = \left\{k : x_i^{k+1}/x_i^k > 1\right\} \quad \text{and} \quad J_-^i = \left\{k : x_i^{k+1}/x_i^k \le 1\right\}.$$

Since $\varphi'(1) = 0$ and φ' is monotone increasing on its domain, we have from (25) that

$$g_i^{k+1} \le 0 \quad \text{for } \forall k \in J_+^i, \quad \forall i \in J(\bar{x}).$$

On the other hand, using (25) and the fact that $\varphi \in \Phi_1 \subseteq \Phi_2$ yields that

$$g_i^{k+1} \ge -\mu\varphi''(1)\lambda_k x_i^k\left(\frac{x_i^{k+1}}{x_i^k} - 1\right) \ge -\mu\varphi''(1)\bar{\lambda}(x_i^{k+1} - x_i^k), \quad \forall k \in J_+^i.$$

Noting that $\lim_{k\to+\infty}(x_i^{k+1} - x_i^k) = 0$, the last two equations imply that

$$\lim_{k\to+\infty,\ k\in J_+^i} g_i^{k+1} = 0, \quad \forall i \in J(\bar{x}). \tag{27}$$

Furthermore, since $\varphi'(t) \le 0$ for any $0 < t \le 1$ by (ii) and (iv), we have from (25) that

$$g_i^{k+1} \ge 0, \quad \forall k \in J_-^i, \quad \forall i \in J(\bar{x}). \tag{28}$$

The inequalities (26)–(28) immediately imply the first part of (24), i.e.,

$$\liminf_{k\to+\infty} g_i^k \ge 0, \quad \forall i = 1, 2, \ldots, n.$$

Next, let us prove the second part of (24). Using (26) and (27) and the fact that $\{x^k\}_{k\in N}$ converges to $\bar{x}$, we have only to prove that

$$\lim_{k\to+\infty,\ k\in J_-^i} g_i^{k+1} x_i^{k+1} = 0, \quad \forall i \in J(\bar{x}).$$

Considering that

$$\lim_{k\to+\infty,\ k\in J_-^i} x_i^{k+1} = 0, \quad \forall i \in J(\bar{x})$$

and using the first part of (24), we then have only to prove that the subsequence $\{g_i^k\}_{k\in J_-^i}$ for each $i \in J(\bar{x})$ is bounded above. Take $\epsilon_0 > 0$ and $x \in \mathbb{R}_{++}^n \cap \operatorname{dom} f$ with $x_i > \epsilon_0$ for any i. Then for $k \in J_-^i$ sufficiently large, we have

$$x_i - x_i^{k+1} \ge \frac{\epsilon_0}{2}, \quad \forall i \in J(\bar{x}). \tag{29}$$

From Definition 2.2, we have

$$f(x) \ge f(x^{k+1}) + \sum_{i=1}^{n} g_i^{k+1}(x - x_i^{k+1}) + o(\|x - x^{k+1}\|), \tag{30}$$

which implies that the subsequence $\{g_i^k\}_{k\in J_-^i}$ is bounded above for $i \in J(\bar{x})$. Indeed, suppose the contrary. Then there would exist an $i_0 \in J(\bar{x})$ and a subsequence $\{g_{i_0}^{k_l}\}_{k_l\in J_-^i}$ (with $\lim_{l\to+\infty} k_l = +\infty$) such that

$$\lim_{l\to+\infty} g_{i_0}^{k_l+1} = +\infty, \quad g_{i_0}^{k_l+1} \ge 0.$$

Since the sequence $\{x^k\}_{k\in N}$ is convergent, using the Eq. 27–29 gives that there exists $\eta \in \mathbb{R}$ such that for sufficiently large l,

$$\sum_{i\ne i_0} g_i^{k_l+1}(x_i - x_i^{k_l+1}) + o(\|x - x^{k+1}\|) \ge \eta.$$

Then, from (29) and (30), we obtain

$$f(x) \ge f(x^{k_l+1}) + \frac{\epsilon_0}{2} g_{i_0}^{k_l+1} + \eta.$$

Since $\lim_{l\to+\infty} f(x^{k_l+1}) \ge f(\bar{x})$ and $\lim_{l\to+\infty} g_{i_0}^{k_l+1} = +\infty$, passing to the limit in the above inequality leads to a contradiction.

(b) From the inequality in (19) and the non-negativity of d_ϕ, it follows that

$$f(x^{k+1}) \leq f(x) + \lambda_k d_\phi(x, x^k), \quad \forall x \in \mathbb{R}^n_{++}.$$

Taking the limit $k \to +\infty$ into the inequality and using $\lim_{k\to+\infty} \lambda_k = 0$, Lemma 3.2 (e) and the lower semicontinuity of f, we then obtain that

$$f(\bar{x}) \leq f(x), \quad \forall x \in \mathbb{R}^n_{++}, \tag{31}$$

where $\bar{x}$ is such that $\lim_{k\to+\infty} x^k = \bar{x}$. This implies that $\bar{x} \in \mathcal{X}^*$. □

Proposition 4.5 *Suppose that assumptions* (A1) *and* (A2) *are satisfied. Let* $\{\lambda_k\}_{k\in N}$ *be an arbitrary sequence of positive numbers and* $\{x^k\}_{k\in N}$ *be the sequence generated by RIPM. Then, the sequence* $\{x^k\}_{k\in N}$ *converges, and furthermore,*

(a) *If there exist* $\underline{\lambda}$ *and* $\bar{\lambda}$ *such that* $0 < \underline{\lambda} < \lambda_k \leq \bar{\lambda}$ *for any k, we have*

$$\liminf_{k\to+\infty} g_i^k \geq 0, \quad \lim_{k\to+\infty} g_i^k x_i^k = 0, \quad \forall i = 1, 2, \ldots, n, \tag{32}$$

where g_i^k *is same as Proposition* 4.4.

(b) *If* $\lim_{k\to+\infty} \lambda_k = 0$*, then the sequence* $\{x^k\}_{k\in N}$ *converges to a solution of problem* (5).

Proof First, we prove that the sequence $\{x^k\}_{k\in N}$ converges. By Lemma 4.3 (b), $\{x^k\}_{k\in N}$ is Fejér convergent to the set U, which implies that

$$\{x^k\}_{k\in N} \subseteq \left\{ y \in \mathbb{R}^n \mid \|x - y\| \leq \|x - x^0\| \right\} \text{ for } \forall x \in U.$$

Note that the latter set is bounded for any given $x \in \mathbb{R}^n$, and therefore, the sequence $\{x^k\}_{k\in N}$ is bounded and there exist an $\hat{x}$ and a subsequence $\{x^{k_j}\}$ of $\{x^k\}_{k\in N}$ converging to $\hat{x}$. Using a similar argument to the first part of Proposition 4.4, we can prove that $\hat{x} \in U$. Thus, by Lemma 4.3 (c), the sequence $\{\|x^k - \hat{x}\|\}_{k\in N}$ is convergent. Since $\{x^{k_j}\} \in \mathbb{R}^n_{++}$ converges to $\hat{x} \in \mathbb{R}^n_+$, we have that $\|\hat{x} - x^{k_j}\| \to 0$, and consequently $\|\hat{x} - x^k\| \to 0$, which implies that the limit point is unique and $x^k \to \hat{x}$.

(a) From the iterative formula (6) and Lemma 2.3 (d), we have

$$0 \in \hat{\partial}\left(f(x) + \lambda_k d_\phi(x, x^k)\right)(x^{k+1}).$$

Therefore, there exists $g^{k+1} \in \hat{\partial} f(x^{k+1})$ such that

$$\lambda_k \nabla_1 d_\phi(x^{k+1}, x^k) = -g^{k+1},$$

i.e., for each $i = 1, 2, \ldots, n$,

$$g_i^{k+1} = -\mu\lambda_k x_i^k \varphi'(x_i^{k+1}/x_i^k) - \nu\lambda_k(x_i^{k+1} - x_i^k). \tag{33}$$

Since $\varphi \in \Phi_2$, we have from (9) that for each $i = 1, 2, \ldots, n$,

$$-\mu\lambda_k x_i^k \varphi'\left(\frac{x_i^{k+1}}{x_i^k}\right) \geq -\mu\lambda_k\varphi''(1)x_i^k\left(\frac{x_i^{k+1}}{x_i^k} - 1\right) \geq -\mu\lambda_k\varphi''(1)(x_i^{k+1} - x_i^k).$$

Combining the last two inequalities then yields that

$$g_i^{k+1} \geq \lambda_k(\mu\varphi''(1) + \nu)(x_i^k - x_i^{k+1}) = 2\theta\lambda_k(x_i^k - x_i^{k+1}), \quad i = 1, 2, \ldots, n, \tag{34}$$

which, together with the facts that $\{x^k\}_{k\in N}$ is convergent and λ_k is bounded, implies the first part of conclusions. In addition, from $\varphi \in \Phi_2$, we have

$$-\mu\lambda_k x_i^k \varphi'\left(\frac{x_i^{k+1}}{x_i^k}\right) \le \mu\lambda_k\varphi''(1)(x_i^k - x_i^{k+1})\frac{x_i^k}{x_i^{k+1}}, \quad i = 1, 2, \ldots, n,$$

which, together with (33), implies that

$$g_i^{k+1} \le \mu\lambda_k\varphi''(1)(x_i^k - x_i^{k+1})\frac{x_i^k}{x_i^{k+1}} + \nu\lambda_k(x_i^k - x_i^{k+1}), \quad i = 1, 2, \ldots, n. \tag{35}$$

Combining the inequalities (34) and (35) yields

$$2\theta\lambda_k x_i^{k+1}(x_i^k - x_i^{k+1}) \le g_i^{k+1}x_i^{k+1} \le (\mu\lambda_k\varphi''(1)x_i^k + \nu\lambda_k x_i^{k+1})(x_i^k - x_i^{k+1}).$$

Considering that $\{x^k\}_{k\in N}$ is convergent, we readily obtain the second part of conclusion from the above inequality. Thus, the proof is completed.
(b) The proof is similar to that of Proposition 4.4 (b), and we here omit it. □

As immediate particular cases of the above propositions, we have the following results.

Corollary 4.1 *Suppose that f is a continuously differentiable quasi-convex function on $\mathbb{R}^n$ and assumptions* (A1) *and* (A2) *are satisfied. Let $\{x^k\}_{k\in N}$ be the sequence generated by IPM or RIPM. Then,*

(a) *if there exist $\underline{\lambda}$ and $\bar{\lambda}$ such that $0 < \underline{\lambda} < \lambda_k \le \bar{\lambda}$ for any k, then $\{x^k\}_{k\in N}$ converges to a stationary point of problem* (5).
(b) *If $\lim_{k\to+\infty}\lambda_k = 0$, then the sequence $\{x^k\}_{k\in N}$ converges to a solution of problem* (5).

Proof (a) From Propositions 4.4–4.5, the sequence $\{x^k\}_{k\in N}$ is convergent and we denote its limit by $\bar{x}$. Then, using Lemma 2.3 and (24) (or (32)) gives

$$(\nabla f(\bar{x}))_i\bar{x}_i = \lim_{k\to+\infty}(\nabla f(x^k))_i x_i^k = 0, \quad i = 1, 2, \ldots, n.$$

This implies that $\nabla f(\bar{x})^T(x - \bar{x}) \ge 0$ for any $x \ge 0$. Therefore, $\bar{x}$ is a stationary point of problem (5) by Definition 2.4 (b).
(b) The proof is similar to that of Proposition 4.4 (b) and Proposition 4.5 (b). □

Particularly, for the limit of the sequence $\{x^k\}_{k\in N}$ generated by RIPM, we have the similar localization result to that of [17]. To see this, let $\pi_U : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$\pi_U(x^0) = \text{argmin}\{\|x - x^0\| : x \in U\}$$

and define

$$\rho(x^0, U) := \|\pi_U(x^0) - x^0\|.$$

Note that $\pi_U(x^0)$ exists since U is a nonempty closed convex set under assumption (A1).

Proposition 4.6 *Suppose that assumptions* (A1) *and* (A2) *are satisfied. Let $\{\lambda_k\}_{k\in N}$ be an arbitrary sequence of positive numbers and $\{x^k\}_{k\in N}$ be the sequence generated by RIPM with the limit $\bar{x}$. Then,*

$$\|x^0 - \bar{x}\| \le 2\rho(x^0, U).$$

Proof Setting $x = \pi_U(x^0)$ in (23), we obtain $\|\pi_U(x^0) - x^k\| \le \|\pi_U(x^0) - x^0\|$. Taking the limit $k \to +\infty$ to this inequality gives

$$\|\pi_U(x^0) - \bar{x}\| \le \|\pi_U(x^0) - x^0\|.$$

By this, we then have

$$\begin{aligned}\|x^0 - \bar{x}\|^2 &\le \left(\|x^0 - \pi_U(x^0)\| + \|\pi_U(x^0) - \bar{x}\|\right)^2 \\ &\le 2\left(\|x^0 - \pi_U(x^0)\|^2 + \|\pi_U(x^0) - \bar{x}\|^2\right) \\ &\le 2\left(\|x^0 - \pi_U(x^0)\|^2 + \|\pi_U(x^0) - x^0\|^2\right) \\ &\le 4\|x^0 - \pi_U(x^0)\|^2 = 4\rho^2(x^0, U).\end{aligned}$$

Thus, the proof is completed. □

5 Numerical experiments

In this section, we report our preliminary numerical experience for IPM and RIPM by solving a class of continuously differentiable quasi-convex optimization problems. We will use the following approximate version of the iterative scheme (6).

AIPM (or ARIPM)

Given $\tau_1, \tau_2 > 0$, and select a starting point $x^0 \in \mathbb{R}^n_{++}$. Set $k := 1$.
For $k = 1, 2, \cdots$, until $|\nabla f(x^k)^T x^k| \le \tau_1$ do

1. Use a unconstrained minimization method to solve approximately the problem

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) + \lambda_k d_\phi(x, x^k) \right\} \tag{36}$$

   and obtain an x^{k+1} such that $\|\nabla f(x^{k+1}) + \lambda_k \nabla_1 d_\phi(x^{k+1}, x^k)\| \le \tau_2$.
2. Let $k := k + 1$, and then go back to Step 1.

End

Unless otherwise stated, $d_\phi(\cdot, \cdot)$ in AIPM and ARIPM is defined as in (4) with

$$\phi(t) = \mu\varphi_2(t) + \frac{\nu}{2}(t-1)^2 = \mu(-\ln t + t - 1) + \frac{\nu}{2}(t-1)^2,$$

where $\mu = 1$ and $\nu = 0$ are used for AIPM whereas $\mu = 1$ and $\nu = 2$ for ARIPM.

We generate a continuously differentiable quasi-convex function $f(x)$ by compounding a quadratic convex function $g(x) = \frac{1}{2}x^T Mx$ with a continuously differentiable increasing function $h\colon I \to \mathbb{R}$, i.e., $f(x) = h(g(x))$, where $I \supseteq \mathbb{R}_+$ and $M \in \mathbb{R}^{n \times n}$ is a given symmetric positive semidefinite matrix. It is not hard to verify that the problem (5) with such $f(x)$ as the objective has a global optimal solution $x^* = 0$. In our experiments, the matrix M was obtained by setting $M = NN^T$, where N is a square matrix whose nonzero elements are chosen randomly from the interval $[-1, 1]$. In this procedure, the number of nonzero elements of N is determined so that the nonzero density of M can be approximately estimated. The function h is given as below.

Experiment A Take $h(t) = -\dfrac{1}{1+t}$ $(t \ne -1)$ and generate five matrices M of dimension $n = 100$ with approximate nonzero density 0.1, 1 and 50%, respectively. Then, we solve each quasi-convex programming problem of form (5) with $f(x) = -\dfrac{1}{1+(x^T Mx)/2}$ by using AIPM and ARIPM.

Experiment B Set $h(t) = \ln(1+t)$ $(t > -1)$ and use those matrices M generated in Experiment A to give the function $f(x) = \ln[1 + (x^T Mx)/2]$. Then, we solve each quasi-convex optimization problem of form (5) with AIPM and ARIPM.

Experiment C Take $h(t) = \arctan(t) + t + 2$ and use those matrices M generated in Experiment A to yield the function $f(x) = \arctan[(x^T Mx)/2] + (x^T Mx)/2 + 2$. Then, we solve each quasi-convex programming problem with AIPM and ARIPM.

Experiment D Take $h(t) = t - \cos(t)$ and use those matrices M in Experiment A to give the corresponding $f(x) = (x^T Mx)/2 - \cos[(x^T Mx)/2]$. Then, we solve each quasi-convex optimization problem with AIPM and ARIPM.

We implemented AIPM and ARIPM with our code in MATLAB 6.5. All numerical experiments were done at a PC with 2.8 GHz CPU and 512 MB memory. We chose a BFGS algorithm with Armijo line search to solve the subproblem (36). To improve numerical behavior, we replaced the standard line search by a nonmonotone line search [9] to seek a suitable steplength, i.e., we computed the smallest non-negative integer l such that

$$F_k(x^k + \beta^l d^k) \le \mathcal{W}_k - \sigma\beta^l F_k(x^k),$$

where $F_k(x) = f(x) + \lambda_k d_\phi(x, x^k)$ and $\mathcal{W}_k$ given by

$$\mathcal{W}_k = \max_{j=k-m_k,\ldots,k} F_k(x^j)$$

and where, for a given non-negative integer $\hat{m}$ and s, we set

$$m_k = \begin{cases} 0, & \text{if } k \le s, \\ \min\left\{m_{k-1} + 1, \hat{m}\right\}, & \text{otherwise.} \end{cases}$$

Throughout the experiments, we used the following parameters for the line search

$$\beta = 0.2, \quad \sigma = 1.0e-4, \quad \hat{m} = 5 \quad \text{and} \quad s = 5.$$

The parameters involved in AIPM and ARIPM, unless otherwise stated, are chosen as

$$x^0 = \omega, \quad \tau_1 = \tau_2 = 1.0e-5, \quad \lambda_k = 1.0e-4 \quad \text{for all } k,$$

where ω is chosen randomly from the interval in [1, 2]. Of course, we should point out that these parameters are not best for all the above experiments.

Numerical results for Experiments A–D are summarized in Tables 1–4. In these tables, **Obj**. denotes the value of $f(x)$ at the final iteration, **Nf** denotes the total number of function evaluations for the objective function $F_k(x)$ of subproblem (36) for solving each quasi-convex programming problem, **Den** denotes the approximate nonzero density of M, and **Time** represents the CPU time for solving each quasi-convex problem.

From Tables 1–4, we see that AIPM and ARIPM can find a favorable approximate optimal solution for all quasi-convex programming problems in Experiments A–C from the given starting point $x^0 = \omega$. However, for the problems in Experiment D where the matrix M has 1 or 50% nonzero density, we adopt the starting point $x^0 = 0.1\omega$ rather than $x^0 = \omega$ so as to obtain favorable results. From these tables, we also observe that, both AIPM and ARIPM will need more function evaluations when the nonzero density of the matrix M becomes higher. In particular, for those problems where M has 50% nonzero density, we must make preprocessing for M by use of scaling technique to get the favorable solution. In addition, we also note that the choice of λ_k in AIPM and ARIPM has an important influence on

Table 1 Numerical results for Experiment A

No.	Den%	IPM			RIPM		
		Obj.	Nf	Time	Obj.	Nf	Time
1	0.1	$-9.999967e-1$	386	0.26	$-9.999962e-1$	365	0.25
2	0.1	$-9.999954e-1$	340	0.17	$-9.999952e-1$	319	0.15
3	0.1	$-9.999999e-1$	477	0.28	$-9.999999e-1$	500	0.28
4	0.1	$-9.999952e-1$	265	0.14	$-9.999964e-1$	260	0.14
5	0.1	$-9.999993e-1$	520	0.30	$-9.999983e-1$	517	0.30
6	1	$-9.999951e-1$	20,617	8.01	$-9.999962e-1$	1,708	1.00
7	1	$-9.999950e-1$	25,531	10.40	$-9.999952e-1$	4,826	2.46
8	1	$-9.999956e-1$	25,397	11.60	$-9.999952e-1$	2,085	0.98
9	1	$-9.999959e-1$	17,891	8.46	$-9.999959e-1$	1,466	0.96
10	1	$-9.999701e-1$	22,905	2.07	$-9.999721e-1$	1,519	0.98
11	50	$-9.999950e-1$	20,583	13.12	$-9.999951e-1$	12,062	7.98
12	50	$-9.999953e-1$	6,531	4.21	$-9.999950e-1$	15,227	8.73
13	50	$-9.999951e-1$	10,074	8.60	$-9.999951e-1$	5,974	4.56
14	50	$-9.999950e-1$	5,246	3.71	$-9.999954e-1$	6,481	6.03
15	50	$-9.999950e-1$	11,120	7.86	$-9.999951e-1$	6,124	4.14

Table 2 Numerical results for Experiment B

No.	Den%	IPM			RIPM		
		Obj.	Nf	Time	Obj.	Nf	Time
1	0.1	$4.013010e-6$	275	0.18	$3.195084e-6$	276	0.18
2	0.1	$4.608092e-6$	303	0.18	$4.800069e-6$	265	0.18
3	0.1	$2.510805e-7$	303	0.18	$3.746151e-6$	298	0.17
4	0.1	$2.842520e-6$	430	0.20	$3.134997e-6$	424	0.21
5	0.1	$5.042468e-7$	303	0.17	$4.966536e-6$	312	0.20
6	1	$3.492764e-6$	1,849	1.15	$4.698351e-6$	1,771	1.15
7	1	$5.042944e-6$	4,663	2.45	$4.863436e-6$	4,382	2.23
8	1	$3.943078e-6$	1,583	1.04	$4.196487e-6$	1,397	0.90
9	1	$3.449366e-6$	1,954	1.31	$4.379457e-6$	1,784	1.20
10	1	$3.329399e-5$	5,732	2.92	$2.281747e-5$	12,485	5.21
11	50	$4.885931e-6$	8,759	7.59	$4.775461e-6$	82,193	77.07
12	50	$4.989218e-6$	17,969	16.20	$4.906823e-6$	18,086	13.54
13	50	$4.988505e-6$	9,215	8.65	$4.827482e-6$	8,817	8.75
14	50	$4.991178e-6$	8,402	8.58	$4.753145e-6$	7,828	7.53
15	50	$4.923310e-6$	9,842	9.25	$4.982674e-6$	9,431	10.21

numerical results. In our future research works, we will make further study for the dynamic choice of λ_k.

6 Conclusions

In this paper, we investigated two classes of entropy-like proximal algorithms based on a second-order homogeneous distance function, namely IPM and RIPM, for the quasi-convex optimization problem (5). The convergence properties of the algorithms were established under some mild assumptions, and particularly, two important convergence results were obtained by imposing certain conditions on the proximal parameters λ_k; see Propositions 4.4–4.5 (a) and (b). Numerical results were reported for those problems where $f(x)$ is a

Table 3 Numerical results for Experiment C

No.	Den%	IPM			RIPM		
		Obj.	Nf	Time	Obj.	Nf	Time
1	0.1	$2.000004e-0$	165	0.14	$2.000004e-0$	169	0.14
2	0.1	$2.000004e-0$	141	0.09	$2.000005e-0$	140	0.07
3	0.1	$2.000004e-0$	240	0.17	$2.000001e-0$	247	0.15
4	0.1	$2.000004e-0$	157	0.09	$2.000002e-0$	157	0.09
5	0.1	$2.000000e-0$	234	1.15	$2.000000e-0$	235	1.14
6	1	$2.000004e-0$	1,426	0.92	$2.000005e-0$	2,038	1.10
7	1	$2.000005e-0$	1,502	0.89	$2.000005e-0$	3,956	2.06
8	1	$2.000005e-0$	1,509	1.12	$2.000005e-0$	1,335	0.89
9	1	$2.000005e-0$	1,517	1.01	$2.000005e-0$	1,572	1.00
10	1	$2.000034e-0$	69,734	19.32	$2.000045e-0$	4,957	2.59
11	50	$2.000005e-0$	3,450	3.43	$2.000005e-0$	3,348	3.51
12	50	$2.000005e-0$	3,759	3.62	$2.000005e-0$	3,689	4.84
13	50	$2.000005e-0$	3,924	3.85	$2.000005e-0$	4,003	4.01
14	50	$2.000005e-0$	3,530	3.96	$2.000005e-0$	3,708	3.71
15	50	$2.000005e-0$	3,963	3.87	$2.000005e-0$	3,940	5.37

Table 4 Numerical results for Experiment D

No.	Den%	IPM			RIPM		
		Obj.	Nf	Time	Obj.	Nf	Time
1	0.1	$9.999984e-1$	557	0.36	$9.999989e-1$	559	0.37
2	0.1	$9.999959e-1$	201	0.15	$9.999961e-1$	449	0.28
3	0.1	$9.999979e-1$	596	0.32	$9.999996e-1$	587	0.31
4	0.1	$9.999954e-1$	416	0.26	$9.999985e-1$	409	0.25
5	0.1	$9.999848e-1$	5,251	1.46	$9.999942e-1$	1,462	0.67
6	1	$9.999954e-1$	1,401	1.06	$9.999951e-1$	1,302	0.98
7	1	$9.999950e-1$	1,596	1.18	$9.999951e-1$	4,444	2.65
8	1	$9.999953e-1$	1,121	0.84	$9.999956e-1$	1,282	0.86
9	1	$9.999968e-1$	1,336	0.96	$9.999950e-1$	1,474	1.00
10	1	$9.999976e-1$	1,521	1.11	$9.999958e-1$	1,600	1.04
11	50	$9.999952e-1$	35,864	24.89*	$9.999951e-1$	4,875	5.05*
12	50	$9.999959e-1$	4,853	6.98	$9.999956e-1$	4,146	3.95
13	50	$9.999951e-1$	4,189	4.14	$9.999951e-1$	4,661	4.96
14	50	$9.999952e-1$	14,078	18.62	$9.999957e-1$	24,907	19.89
15	50	$9.999961e-1$	3,577	3.90	$9.999974e-1$	4,155	4.12
15	50	$9.995090e-1$	2,128	1.82	$9.999950e-1$	35,463	33.26

For the problems where the matrix M has 1 or 50% nonzero density, the starting point $x^0 = 0.1\omega$ was used in the experiments. The notation "$*$" means that the results were obtained by using $\lambda_k = 1.0e-3$ and $1.0e-5$ for IPM and RIPM, respectively

composition of a quadratic convex function and a continuously differentiable increasing function, which indicate that IPM and RIPM are effective for this class of quasi-convex programming problems, and they are very promising for finding the global optimal solution. All convergence results of IPM and RIPM in this paper can be easily extended to the following more general quasi-convex problem

$$\begin{aligned} &\min \ f(x) \\ &\text{s.t.} \quad Ax + b \geq 0, \end{aligned}$$

where the matrix $A \in \mathbb{R}^{m\times n}$ has a full column rank and $b \in \mathbb{R}^m$.

As pointed out by a referee, the sets Φ_1 and Φ_2 are two convex cones. This implies that for any $\varphi_1, \varphi_2 \in \Phi_2$ and any $\alpha, \beta > 0$, we have

$$\alpha\varphi_1 + \beta\varphi_2 \in \Phi_2.$$

For example, from the first two examples given in Sect. 3, we have

$$\psi(t) = \varphi_1(t) + \varphi_2(t) = (t \ln t - t + 1) + (-\ln t + t - 1) = (t - 1)\ln t \in \Phi_2.$$

Similarly, there are a lot of concrete functions belong to Φ_2, which can be used to design proximal algorithms for the quasi-convex optimization problem. Thus, an interesting problem will arise: what kind of functions like this have better numerical performance. In our future research work, we will make study for the comparison of these kernel functions.

Acknowledgment The authors would like to thank two anonymous referees for their helpful comments and suggestions on the revision of this paper.

References

1. Auslender, A., Teboulle, M., Ben-TBA, S.: A logarithmic-quadratic proximal method for variational inequalities. Comput. Optim. Appl. **12**, 31–40 (1999)
2. Auslender, A., Teboulle, M., Ben-TIBA, S.: Interior proximal and multiplier methods based on second order homogeneous kernels. Math. Oper. Res. **24**, 645–668 (1999)
3. Bazaraa, M.S., Shetty, C.M.: Nonlinear Programming Theory and Algorithms. Wiley, New York (1979)
4. Censor, Y., Zenios, S.A.: The proximal minimization algorithm with D-functions. J. Optim. Theory Appl. **73**, 451–464 (1992)
5. Chen, G., Teboulle, M.: Convergence analysis of a proximal-like minimization algorithm using Bregman functions. SIAM J. Optim. **3**, 538–543 (1993)
6. Doljansky, M., Teboulle, M.: An interior proximal algorithm and the exponential multiplier method for semidefinite programming. SIAM J. Optim. **9**, 1–13 (1998)
7. Eckstein, J.: Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. Math. Oper. Res. **18**, 206–226 (1993)
8. Eggermont, P.B.: Multilicative iterative algorithms for convex programming. Linear Algebra Appl. **130**, 25–42 (1990)
9. Grippo, L., Lampariello, F., Lucidi, S.: A nonmonotone line search technique for Newton's method. SIAM J. Numer. Anal. **23**, 707–716 (1986)
10. Guler, O.: On the convergence of the proximal point algorithm for convex minimization. SIAM J. Control Optim. **29**, 403–419 (1991)
11. Horst, R., Tuy, H.: Global Optimization (Deterministic Approach), 3rd edn. Springer, Berlin (1996)
12. Iusem, A., Svaiter, B., Teboulle, M.: Entropy-Like proximal methods in convex programming. Math. Oper. Res. **19**, 790–814 (1994)
13. Iusem, A., Teboulle, M.: Convergence rate analysis of nonquadratic proximal method for convex and linear programming. Math. Oper. Res. **20**, 657–677 (1995)
14. Kiwiel, K.C.: Proximal minimization methods with generalized Bregman functions. SIAM J. Control Optim. **35**, 1142–1168 (1997)
15. Martinet, B.: Perturbation des methodes d'Optimisation. Appl. R.A.I.R.O. Numer. Anal. **12**, 154–171 (1978)
16. Moreau, J.J.: Promimité et Dualité dans un Espace Hilbertien. Bull. Soc. Math. France **93**, 273–299 (1965)
17. Papa Quiroz, E.A., Roberto Qliveira P.: Convergence of the proximal point method for quasiconvex minimization. Technical Report, Federal University of Rio de Janeiro, Brazil (2006)
18. Rockafella, R.T.: Convex Analysis. Princeton University Press, Princeton, NJ (1970)
19. Rockafellar, R.T.: Augmented lagrangians and applications of proximal point algorithm in convex programming. Math. Oper. Res. **1**, 97–116 (1976)
20. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–898 (1976)
21. Rockafellar, R.T., Wets, R.: Variational Analysis. Springer, Berlin (1998)

22. Tseng, P., Bertsekas, D.P.: On the convergence of the exponential multiplier method for convex programming. Math. Program **60**, 1–19 (1993)
23. Teboulle, M.: Entropic proximal mappings with applications to nonlinear programming. Math. Oper. Res. **17**, 670–690 (1992)
24. Teboulle, M.: Convergence of proximal-like algorithms. SIAM J. Optim. **7**, 1069–1083 (1997)