ACCELERATED MIRROR DESCENT METHOD THROUGH VARIABLE AND OPERATOR SPLITTING

Anonymous authors

Paper under double-blind review

ABSTRACT

Accelerated Mirror Descent (Acc-MD) is derived from a discretization of an accelerated mirror ODE system using a variable–operator splitting framework. A new Cauchy–Schwarz type inequality enables the first proof of linear accelerated convergence for mirror descent on a broad class of problems. Unlike prior methods based on the triangle scaling exponent (TSE), Acc-MD achieves acceleration in some cases where TSE fails. Experiments on smooth and composite optimization tasks show that Acc-MD consistently outperforms existing accelerated variants, both theoretically and empirically.

1 Introduction

Consider the unconstrained problem

$$\min_{x \in \mathbb{R}^n} f(x),$$

where f is convex but may not have a Lipschitz-continuous gradient. Let ϕ be a smooth and strictly convex *mirror function* that defines the underlying geometry. Mirror descent (Nemirovskij and Yudin, 1983) updates as

$$\nabla \phi(x_{k+1}) - \nabla \phi(x_k) = -\alpha_k \nabla f(x_k),$$

extending gradient descent to non-Euclidean settings. Nesterov (2005) proposed an accelerated variant using estimate sequences.

Recent work views optimization algorithms as discretizations of continuous-time dynamics whose trajectories converge to minimizers. Wibisono et al. (2016) derived a Bregman Lagrangian that captures accelerated flows and showed that its discretization gives accelerated mirror descent. Krichene et al. (2015) extended the ODE framework for Nesterov acceleration (Su et al., 2016) to accelerated mirror descent, with discretizations yielding a class of first-order methods with $\mathcal{O}(L_f/k^2)$ convergence, where L_f is the Lipschitz constant of ∇f . More recently, xiang Yuan and Zhang (2023) developed high-resolution ODEs and recovered the optimal $\mathcal{O}(L_f/k^2)$ rate for accelerated mirror descent methods from (Nesterov, 2005).

However, the analyses in (Nesterov, 2005; Wibisono et al., 2016; Krichene et al., 2015; xiang Yuan and Zhang, 2023) do not incorporate the notion of relative smoothness, i.e., $D_f(x,y) \leq LD_\phi(x,y)$, where D_f and D_ϕ denote the Bregman divergence of f and ϕ , respectively. As a result, their convergence guarantees may be loose outside the Euclidean setting, especially when $L \ll L_f$. This reflects a limitation of Nesterov's original theory (Nesterov, 2005), which does not fully use the geometry induced by the mirror map.

Under the relative smoothness condition, mirror descent achieves an $\mathcal{O}(L/k)$ convergence rate, as first shown in (Birnbaum et al., 2011). The notion of relative strong convexity, $\mu D_{\phi}(x,y) \leq D_f(x,y)$, was later introduced in (Lu et al., 2018), leading to a linear convergence rate of $1 - \mu/L$.

Mirror descent also extends to composite optimization problems of the form F(x) = f(x) + g(x), where f is smooth and g is convex but may be non-smooth. The resulting Bregman Proximal Gradient (BPG) method (Teboulle, 2018) achieves a convergence rate of $\mathcal{O}(L/k)$ under relative smoothness; see also (Bauschke et al., 2017; Lu et al., 2018; Zhou et al., 2019).

Hanzely et al. (2021) proposed the Accelerated Bregman Proximal Gradient (ABPG) method, which attains an accelerated rate of $\mathcal{O}(L/k^{\gamma})$ using the *triangle scaling exponent* (TSE) γ . Full acceleration

 $\mathcal{O}(L/k^2)$ requires $\gamma=2$ in both the algorithm and the analysis, which limits applicability and prevents generalization to broader Bregman geometries. In addition, no accelerated linear convergence has been shown under relative strong convexity.

Contributions The main contributions of this work are as follows:

• Building on the variable and operator splitting framework of (Chen et al., 2025), we propose a new accelerated mirror descent (Acc-MD) flow with initial conditions $x(0) = x_0$, $y(0) = y_0$:

$$\begin{cases} x' = y - x, \\ (\nabla \phi(y))' = -\mu^{-1} \nabla f(x) + \nabla \phi(x) - \nabla \phi(y), \end{cases}$$
(1)

We split the variable into x and y. At equilibrium, $y^* = x^*$ and $\nabla f(x^*) = 0$, recovering stationarity. The auxiliary variable y decouples the primal update and inertial motion.

- From an implicit—explicit discretization of (1), we design an Acc-MD method; see Algorithm 1. The method is simpler than existing counterparts in the literature, e.g., algorithms in (Nesterov, 2005; Krichene et al., 2015; Hanzely et al., 2021).
- We give the first accelerated linear convergence guarantee under relative strong convexity μ , with rate $(1 + \sqrt{\mu/C_{f,\phi}})^{-1}$, where $C_{f,\phi}$ comes from a Cauchy–Schwarz inequality; see Assumption (A2). We show that in some cases, even when the TSE parameter $\gamma = 1$, our method achieves acceleration while ABPG (Hanzely et al., 2021) does not.
- We extend the framework to the convex case $\mu = 0$ using a perturbation and homotopy argument, and recover the optimal $\mathcal{O}(C_{f,\phi}/k^2)$ accelerated rate with Assumption (A2). Without (A2), it is shown in (Dragomir et al., 2022) that the optimal complexity of mirror descent methods is $\mathcal{O}(1/k)$.
- We extend the algorithm to composite optimization $\min_x f(x) + g(x)$. With variable and operator splitting $\nabla f(x) + \nabla g(y)$, the change is minimal:

$$(\nabla \phi(y))' = -\mu^{-1}(\nabla f(x) + \nabla g(y)) + \nabla \phi(x) - \nabla \phi(y), \tag{2}$$

and the algorithm follows from implicit discretization in $\nabla g(y)$. The method works for non-smooth g with a proximal operator. Since some constrained problems can be written as composite optimization, Acc-MD also applies to constrained settings.

• We benchmark Acc-MD on a range of smooth, non-smooth, and constrained convex problems, showing consistent advantages in both theory and practice over existing mirror descent methods.

Limitation While our approach avoids the triangle scaling property, both the algorithm and the analysis rely on accurate estimates of the relative strong convexity μ and the constant $C_{f,\phi}$. Developing adaptive schemes for these parameters is an important direction for future work.

Although we provide examples where (A2) holds but TSE fails to give acceleration, we cannot expect (A2) to be verified for every problem. Further relaxation of this assumption deserves study.

Another limitation is the convexity assumption. Extending the analysis to non-convex objectives remains open. In such settings, the design of suitable non-convex mirror functions may be important for guiding convergence, but theoretical guarantees are currently lacking.

Preliminaries Let V be a normed vector space. For $f \in \mathcal{C}^1(V)$, define the Bregman divergence

$$D_f(y,x) := f(y) - f(x) - \langle \nabla f(x), y - x \rangle.$$

For $f \in \mathcal{C}^1(V)$, f is convex if and only if $D_f(y,x) \geq 0$ for all $x,y \in V$. If f is strictly convex, then $D_f(y,x) = 0$ holds if and only if x = y. In general, the Bregman divergence is not symmetric, i.e., $D_f(y,x) \neq D_f(x,y)$. Its symmetrization is

$$D_f(y,x) + D_f(x,y) = \langle \nabla f(y) - \nabla f(x), y - x \rangle.$$

A key tool in the convergence analysis is the three-point identity of Bregman divergence (Chen and Teboulle, 1993), which follows directly from the definition:

$$\langle \nabla f(y) - \nabla f(x), y - z \rangle = D_f(y, x) + D_f(z, y) - D_f(z, x), \tag{3}$$

Let ϕ^* be the convex conjugate of ϕ . The mappings between primal and dual variables are

$$\chi = \nabla \phi(x), \quad x = \nabla \phi^*(\chi), \quad \eta = \nabla \phi(y), \quad y = \nabla \phi^*(\eta).$$

Here (x,y) are the primal variables and (χ,η) are their dual counterparts. The maps $\nabla \phi: V \to V^*$ and $\nabla \phi^*: V^* \to V$ are assumed to be efficiently computable. To clarify, we refer to ϕ as the mirror function and $\nabla \phi$ as the mirror map.

An important symmetry relation connects the Bregman divergences of ϕ and its conjugate:

$$D_{\phi}(x,y) = D_{\phi^*}(\eta,\chi),\tag{4}$$

with reversed argument order. Moreover, the gradient of the Bregman divergence with respect to its first argument satisfies

$$\nabla D_f(\cdot, x) = \nabla f(\cdot) - \nabla f(x), \quad \nabla D_{\phi^*}(\cdot, \chi) = \nabla \phi^*(\cdot) - \nabla \phi^*(\chi). \tag{5}$$

Let A be a self-adjoint, positive definite operator on a Hilbert space V with inner product (\cdot, \cdot) . Then

$$(x,y)_A := (Ax,y)$$

defines a new inner product, and the associated norm is denoted by $\|\cdot\|_A$. The corresponding dual norm is $\|\cdot\|_{A^{-1}}$. The convexity and Lipschitz constants of a differentiable function f relative to $\|\cdot\|_A$ are defined by

$$\mu_f(A) \|x - y\|_A^2 \le \langle \nabla f(x) - \nabla f(y), x - y \rangle \le L_f(A) \|x - y\|_A^2 \quad \forall x, y \in V.$$

Choosing an appropriate inner product may reduce the condition number $\kappa_A(f) := L_f(A)/\mu_f(A)$, in which case A is called a *preconditioner*.

Analogously, the mirror function ϕ may be viewed as a *nonlinear preconditioner*. This is especially useful when L_f is unbounded but the relative smoothness L is finite.

2 MIRROR DESCENT METHODS

Given an initial condition $x(0) = x_0$, the mirror descent flow is defined by

$$\frac{\mathrm{d}}{\mathrm{d}t}\nabla\phi(x(t)) = -\nabla f(x(t)). \tag{6}$$

Discretizing the flow (6) using the explicit Euler method yields the iteration

$$\nabla \phi(x_{k+1}) - \nabla \phi(x_k) = -\alpha_k \nabla f(x_k), \tag{7}$$

where $\alpha_k > 0$ is the step size. This is equivalent to the classical mirror descent update (Beck and Teboulle, 2003; Bubeck, 2015; Lu et al., 2018):

$$x_{k+1} = \arg\min_{x \in V} \left\{ \langle \nabla f(x_k), x \rangle + \frac{1}{\alpha_k} D_{\phi}(x, x_k) \right\}.$$

As noted by Krichene et al. (2015), the flow (6) can also be written in dual form:

$$\chi' = -\nabla f(x), \quad x = \nabla \phi^*(\chi), \tag{8}$$

which is closely related to the Bregman Inverse Scale Space dynamics (Osher et al., 2016).

We present the following three-point identity connecting the Bregman divergences of ϕ and f, which simplifies the convergence analysis compared to prior work, e.g., (Lu et al., 2018).

Lemma 2.1. Let $\{x_k\}$ be the sequence generated by the mirror descent method (7). Then

$$D_{\phi}(x^*, x_{k+1}) - D_{\phi}(x^*, x_k) + D_{\phi}(x_{k+1}, x_k)$$

$$= \alpha_k \Big[-D_f(x_{k+1}, x^*) - D_f(x^*, x_k) + D_f(x_{k+1}, x_k) \Big].$$
(9)

Proof. Using the Bregman identity (3) and the update rule (7), we compute

$$D_{\phi}(x^*, x_{k+1}) - D_{\phi}(x^*, x_k) + D_{\phi}(x_{k+1}, x_k) = \langle \nabla \phi(x_{k+1}) - \nabla \phi(x_k), x_{k+1} - x^* \rangle$$

= $-\alpha_k \langle \nabla f(x_k) - \nabla f(x^*), x_{k+1} - x^* \rangle = \alpha_k \left[-D_f(x_{k+1}, x^*) - D_f(x^*, x_k) + D_f(x_{k+1}, x_k) \right].$

To establish linear convergence, we adopt the notions of relative smoothness and relative strong convexity introduced in (Lu et al., 2018). The significance of relative smoothness is that f may not be smooth, i.e., $L_f = \infty$, while L remains bounded.

(A1) The convex function f is said to be *relatively smooth* and *relatively convex* with respect to a mirror function ϕ if there exist constants $\mu \geq 0$ and $L \geq \mu$ such that

$$\mu D_{\phi}(x,y) \le D_f(x,y) \le LD_{\phi}(x,y) \quad \forall x, y \in V. \tag{10}$$

Using the linearity of the Bregman divergence, this is equivalent to $L\phi - f$ and $f - \mu\phi$ being convex.

Theorem 2.2. Suppose f satisfies assumption (A1), and let $\{x_k\}$ be the sequence generated by the mirror descent iteration (7). Then for any step size $\alpha_k \leq 1/L$, we have the decay property

$$D_{\phi}(x^*, x_k) - D_{\phi}(x^*, x_{k+1}) \ge \alpha_k \left[D_f(x_{k+1}, x^*) + D_f(x^*, x_k) \right]. \tag{11}$$

In particular, if $\alpha_k = 1/L$ for all $k \ge 0$, then linear convergence holds:

$$D_{\phi}(x^*, x_k) \le \left(1 - \frac{\mu}{L}\right)^k D_{\phi}(x^*, x_0).$$
 (12)

Proof. From the identity (9) and assumption (A1), we have

$$D_{\phi}(x^*, x_{k+1}) - D_{\phi}(x^*, x_k) \le -\alpha_k \left[D_f(x_{k+1}, x^*) + D_f(x^*, x_k) \right] + (\alpha_k L - 1) D_{\phi}(x_{k+1}, x_k)$$

$$\le -\alpha_k \left[D_f(x_{k+1}, x^*) + D_f(x^*, x_k) \right] \le -\alpha_k \mu D_{\phi}(x^*, x_k).$$

Choosing $\alpha_k \equiv 1/L$ and rearranging the inequality to get the desired linear convergence.

Similar results hold for proximal mirror descent methods (e.g., BPG) for composite optimization. We omit the details here, as an accelerated version will be presented in the next section.

3 ACCELERATED MIRROR DESCENT METHODS

We follow the recent variable and operator splitting (VOS) framework of (Chen et al., 2025) to develop accelerated mirror descent methods. We assume that f satisfies assumption (A1) with constants L and $\mu > 0$.

Flow Introducing dual variables $\chi = \nabla \phi(x)$ and $\eta = \nabla \phi(y)$, we can write the accelerated mirror descent flow (1) as

$$x' = y - x, \quad \eta' = -\mu^{-1} \nabla f_{-\mu}(x) - \eta,$$
 (13)

where $f_{-\mu} := f - \mu \phi$, or equivalently $f = f_{-\mu} + \mu \phi$. Under assumption (A1), $f_{-\mu}$ is convex.

Stability We define a Lyapunov function:

$$\mathcal{E}(x,\eta) := D_{f_{-\mu}}(x,x^*) + \mu D_{\phi^*}(\eta,\chi^*), \tag{14}$$

where x^* is a minimizer of f and $\chi^* = \nabla \phi(x^*)$. This energy couples primal and dual Bregman divergences, capturing the geometry induced by the mirror map. We establish exponential stability by verifying a strong Lyapunov property (Chen and Luo, 2021).

Lemma 3.1. Let $\mathcal{E}(x,\eta)$ be defined by (14), and define the vector field $\mathcal{G}(x,\eta) = (y-x, -\mu^{-1}\nabla f_{-\mu}(x) - \eta)$, where the dual variables satisfy $\chi = \nabla \phi(x)$ and $\eta = \nabla \phi(y)$. Then

$$-\nabla \mathcal{E}(x,\eta) \cdot \mathcal{G}(x,\eta) = \mathcal{E}(x,\eta) + D_{f_{-\mu}}(x^*,x) + \mu D_{\phi}(y,x^*). \tag{15}$$

As a consequence, any solution (x(t), y(t)) of the flow (1) satisfies the exponential decay bound

$$\mathcal{E}(x(t), \eta(t)) \le e^{-t} \mathcal{E}(x(0), \eta(0)). \tag{16}$$

Proof. Using the identity (5), we compute the gradients of the Lyapunov function:

$$\partial_x \mathcal{E} = \nabla f_{-\mu}(x) - \nabla f_{-\mu}(x^*), \quad \partial_n \mathcal{E} = \mu \left(\nabla \phi^*(\eta) - \nabla \phi^*(\chi^*) \right) = \mu (y - x^*).$$

Thus, by the direct calculation and symmetry relation (4), we have

$$-\nabla \mathcal{E}(x,\eta) \cdot \mathcal{G}(x,\eta) = \langle \nabla f_{-\mu}(x) - \nabla f_{-\mu}(x^*), x - y \rangle + \mu \langle y - x^*, \mu^{-1}(\nabla f_{-\mu}(x) - \nabla f_{-\mu}(x^*)) + \eta - \chi^* \rangle = \langle \nabla f_{-\mu}(x) - \nabla f_{-\mu}(x^*), x - x^* \rangle + \mu \langle \nabla \phi^*(\eta) - \nabla \phi^*(\chi^*), \eta - \chi^* \rangle = D_{f_{-\mu}}(x, x^*) + D_{f_{-\mu}}(x^*, x) + \mu D_{\phi^*}(\chi^*, \eta) + \mu D_{\phi^*}(\eta, \chi^*) = \mathcal{E}(x, \eta) + D_{f_{-\mu}}(x^*, x) + \mu D_{\phi}(y, x^*).$$

Since $D_{f_{-\mu}}(x^*, x) \ge 0$ and $D_{\phi}(y, x^*) \ge 0$ by convexity of $f_{-\mu}$ and ϕ , we obtain the inequality $\nabla \mathcal{E}(x, \eta) \cdot \mathcal{G}(x, \eta) \le -\mathcal{E}(x, \eta)$,

from which the exponential decay (16) follows by Grönwall's inequality.

As a corollary, we have $x(t), y(t) \in B(x^*, R)$ for all $t \ge 0$, where $B(x^*, R)$ denotes the ball of radius R centered at x^* .

Accelerated Mirror Descent Methods To obtain the accelerated convergence rate, we need one more assumption.

(A2) There exists $C_{f,\phi} > 0$ such that the following relative Cauchy Schwarz inequality holds

$$|\langle \nabla f_{-\mu}(x) - \nabla f_{-\mu}(\hat{x}), y - \hat{y} \rangle| \le 2\sqrt{C_{f,\phi}} D_{f_{-\mu}}^{1/2}(x,\hat{x}) D_{\phi}^{1/2}(\hat{y},y), \quad \forall x, \hat{x}, y, \hat{y} \in V.$$

We propose an accelerated mirror descent (Acc-MD) method by an implicit-explicit scheme of (1):

$$\frac{x_{k+1} - x_k}{\alpha} = 2y_{k+1} - y_k - x_{k+1},\tag{17a}$$

$$\frac{\nabla \phi(y_{k+1}) - \nabla \phi(y_k)}{\alpha} = -\frac{1}{\mu} \nabla f(x_k) + \nabla \phi(x_k) - \nabla \phi(y_{k+1}). \tag{17b}$$

Instead of a standard explicit Euler discretization, y is discretized semi-implicitly as $(2y_{k+1}-y_k)$, and x implicitly as x_{k+1} in (17a). In (17b), $\nabla f_{-\mu}(x_k)$ is explicit and $-\nabla \phi(y_{k+1})$ is implicit. The implicit term improves stability, similar to the proximal point algorithm. The discretization $y \approx (2y_{k+1}-y_k)$ acts as an accelerated over-relaxation (AOR), symmetrizing the error equation and enabling our proof of the accelerated rate.

An equivalent computation favorable form is given as Algorithm 1.

Algorithm 1 Accelerated mirror descent (Acc-MD) method

```
1: Parameters: x_0, y_0 \in \mathbb{R}^n, \mu, C_{f,\phi}.
```

2: Set
$$\alpha = \sqrt{\mu/C_{f,\phi}}$$
.

3: **for**
$$k = 0, 1, 2, \dots$$
 do

4:
$$y_{k+1} = \arg\min_{y \in \mathbb{R}^n} (1+\alpha)\phi(y) - \left\langle -\frac{\alpha}{\mu} \nabla f(x_k) + \alpha \nabla \phi(x_k) + \nabla \phi(y_k), y \right\rangle.$$

5:
$$x_{k+1} = \frac{1}{1+\alpha} [x_k + \alpha(2y_{k+1} - y_k)].$$

6: end for

Convergence analysis For four points (x, \hat{x}, y, \hat{y}) and $\alpha \in \mathbb{R}$, introduce the cross term

$$\mathcal{B}^{\alpha}(x,\hat{x},\hat{y},y) = D_{f_{-\mu}}(x,\hat{x}) + \mu D_{\phi}(\hat{y},y) - \alpha \langle \nabla f_{-\mu}(x) - \nabla f_{-\mu}(\hat{x}), y - \hat{y} \rangle. \tag{18}$$

Lemma 3.2. Let Assumption (A2) hold. Then for $|\alpha| \leq \sqrt{\mu/C_{f,\phi}}$, we have $\mathcal{B}^{\alpha}(x,\hat{x},\hat{y},y) \geq 0$.

Proof. By (A2) and the inequality $2ab < a^2 + b^2$, we have

$$\alpha |\langle \nabla f_{-\mu}(x) - \nabla f_{-\mu}(\hat{x}), y - \hat{y} \rangle| \leq 2\alpha \sqrt{C_{f,\phi}} D_{f_{-\mu}}^{1/2}(x, \hat{x}) D_{\phi}^{1/2}(\hat{y}, y)$$
$$\leq \alpha \sqrt{C_{f,\phi}/\mu} \left(D_{f_{-\mu}}(x, \hat{x}) + \mu D_{\phi}(\hat{y}, y) \right).$$

Therefore
$$\mathcal{B}^{\alpha}(x,\hat{x},\hat{y},y) \geq 0$$
 if $\alpha \leq \sqrt{\mu/C_{f,\phi}}$.

Define the modified Lyapunov function

$$\mathcal{E}^{\alpha}(x,y) := \mathcal{B}^{\alpha}(x,x^*,x^*,y) := \mathcal{E}(x,\eta) - \alpha \left\langle \nabla f_{-\mu}(x) - \nabla f_{-\mu}(x^*), y - x^* \right\rangle. \tag{19}$$

As shown in Lemma 3.2, when $\alpha \leq \sqrt{\mu/C_{f,\phi}}$, $\mathcal{E}^{\alpha}(x,y) \geq 0$ is indeed a Lyapunov function.

Lemma 3.3. Let (x_k, y_k) be the sequence generated by Acc-MD iterations (17), then it holds

$$\mathcal{E}^{\alpha}(x_{k+1}, y_{k+1}) - \mathcal{E}^{\alpha}(x_k, y_k) = -\alpha \mathcal{E}^{\alpha}(x_{k+1}, y_{k+1}) - \alpha \mathcal{B}^{-\alpha}(x^*, x_{k+1}, y_{k+1}, x^*) - \mathcal{B}^{\alpha}(x_k, x_{k+1}, y_{k+1}, y_k).$$
(20)

Proof. Denote $z = (x, \eta)$. Expand the difference of the Lyapunov function $\mathcal{E}(z)$ at z_{k+1} :

$$\begin{split} \mathcal{E}(\boldsymbol{z}_{k+1}) - \mathcal{E}(\boldsymbol{z}_{k}) &= \langle \nabla \mathcal{E}(\boldsymbol{z}_{k+1}), \boldsymbol{z}_{k+1} - \boldsymbol{z}_{k} \rangle - D_{\mathcal{E}}(\boldsymbol{z}_{k}, \boldsymbol{z}_{k+1}) \\ &= \alpha \left\langle \nabla \mathcal{E}(\boldsymbol{z}_{k+1}), \mathcal{G}(\boldsymbol{z}_{k+1}) \right\rangle - D_{\mathcal{E}}(\boldsymbol{z}_{k}, \boldsymbol{z}_{k+1}) \\ + \alpha \left\langle \nabla f_{-\mu}(\boldsymbol{x}_{k+1}) - \nabla f_{-\mu}(\boldsymbol{x}^{*}), \boldsymbol{y}_{k+1} - \boldsymbol{y}_{k} \right\rangle + \alpha \left\langle \boldsymbol{y}_{k+1} - \boldsymbol{x}^{*}, \nabla f_{-\mu}(\boldsymbol{x}_{k+1}) - \nabla f_{-\mu}(\boldsymbol{x}_{k}) \right\rangle. \end{split}$$

The last line is the difference with the implicit Euler discretization and can be symmetrized as

$$\alpha \left\langle \nabla f_{-\mu}(x_{k+1}) - \nabla f_{-\mu}(x^*), y_{k+1} - x^* \right\rangle - \alpha \left\langle \nabla f_{-\mu}(x_k) - \nabla f_{-\mu}(x^*), y_k - x^* \right\rangle + \alpha \left\langle \nabla f_{-\mu}(x_{k+1}) - \nabla f_{-\mu}(x_k), y_{k+1} - y_k \right\rangle.$$

Then we use identity (15) to expand $\langle \nabla \mathcal{E}(z_{k+1}), \mathcal{G}(z_{k+1}) \rangle$ and rearrange the terms to get

$$\begin{split} \mathcal{E}^{\alpha}(x_{k+1},y_{k+1}) - \mathcal{E}^{\alpha}(x_{k},y_{k}) &= -\alpha \mathcal{E}(x_{k+1},\eta_{k+1}) - \alpha D_{f_{-\mu}}(x^{*},x_{k+1}) - \mu \alpha D_{\phi}(y_{k+1},x^{*}) \\ &- D_{\mathcal{E}}(\boldsymbol{z}_{k},\boldsymbol{z}_{k+1}) + \alpha \left\langle \nabla f_{-\mu}(x_{k+1}) - \nabla f_{-\mu}(x_{k}),y_{k+1} - y_{k} \right\rangle \\ &= -\alpha \mathcal{E}^{\alpha}(x_{k+1},y_{k+1}) - \alpha D_{f_{-\mu}}(x^{*},x_{k+1}) - \mu \alpha D_{\phi}(y_{k+1},x^{*}) \\ &- \alpha \left\langle \nabla f_{-\mu}(x_{k+1}) - \nabla f_{-\mu}(x^{*}),y_{k+1} - x^{*} \right\rangle \\ &- \mathcal{B}^{\alpha}(x_{k},x_{k+1},y_{k+1},y_{k}). \end{split}$$

Theorem 3.4 (Convergence of Acc-MD method). Suppose f is μ -relatively convex with $\mu > 0$ and (A2) holds with constant $C_{f,\phi}$. Let (x_k,y_k) be generated by scheme (17) with initial value (x_0,y_0) , $\eta_k = \nabla \phi(y_k)$, and step size $\alpha = \sqrt{\mu/C_{f,\phi}}$. Then there exists a constant $C_0 = C_0(x_0,y_0,\mu,C_{f,\phi})$ so that we have the accelerated linear convergence

$$D_{f_{-\mu}}(x_{k+1}, x^*) + \mu D_{\phi^*}(\eta_{k+1}, \chi^*) \le C_0 \left(\frac{1}{1 + \sqrt{\mu/C_{f,\phi}}}\right)^k, \quad k \ge 1.$$
 (21)

Proof. By Lemma 3.2, for $\alpha = \sqrt{\mu/C_{f,\phi}}$, we can drop negative terms from the identity (20) to get the linear convergence

$$\mathcal{E}^{\alpha}(x_{k+1}, y_{k+1}) \le \frac{1}{1+\alpha} \mathcal{E}^{\alpha}(x_k, y_k) \le \left(\frac{1}{1+\sqrt{\mu/C_{f,\phi}}}\right)^{k+1} \mathcal{E}^{\alpha}(x_0, y_0). \tag{22}$$

From the proof of Lemma 3.3, we have

$$\alpha \mathcal{E}(x_{k+1}, \eta_{k+1}) \le \mathcal{E}^{\alpha}(x_k, y_k) - \mathcal{E}^{\alpha}(x_{k+1}, y_{k+1}) \le \left(\frac{1}{1 + \sqrt{\mu/C_{f,\phi}}}\right)^k \mathcal{E}^{\alpha}(x_0, y_0),$$

which leads to (21).

Discussion on (A2) Without **(A2)**, it is shown in (Dragomir et al., 2022) that the optimal complexity of mirror descent type methods is $\mathcal{O}(1/k)$. Hence, **(A2)** is crucial to achieve acceleration. We now verify Assumption **(A2)** for a large class of functions.

Theorem 3.5. Assume $f_{-\mu} = f - \mu \phi$ is smooth and ϕ is strongly convex. Then

$$C_{f,\phi} \le \inf_{SPD} \frac{L_{f_{-\mu}}(A)}{\mu_{\phi}(A)} \le \frac{L_{f_{-\mu}}}{\mu_{\phi}}.$$

Proof. For any SPD A, we have

$$\langle \nabla f_{-\mu}(x) - \nabla f_{-\mu}(\hat{x}), y - \hat{y} \rangle = \langle A^{-1/2}(\nabla f_{-\mu}(x) - \nabla f_{-\mu}(\hat{x})), A^{1/2}(y - \hat{y}) \rangle$$

$$\leq \|\nabla f_{-\mu}(x) - \nabla f_{-\mu}(\hat{x})\|_{A^{-1}} \|y - \hat{y}\|_{A}.$$

Applying strong convexity of ϕ in $\|\cdot\|_A$ norm and co-coercivity of $f_{-\mu}$ in $\|\cdot\|_{A^{-1}}$ norm, we obtain

$$2D_{\phi}(\hat{y}, y) \ge \mu_{\phi}(A) \|y - \hat{y}\|_{A}^{2}, \quad 2D_{f_{-\mu}}(x, \hat{x}) \ge \frac{1}{L_{f_{-\mu}}(A)} \|\nabla f_{-\mu}(x) - \nabla f_{-\mu}(\hat{x})\|_{A^{-1}}^{2}.$$

Combining the above yields the first estimate. Taking A = I to get the second.

The mirror functions ϕ can be rescaled so that $\mu_{\phi}=1$. Theorem 3.5 can be used to verify (A2) for two examples in (Collins et al., 2008). We write $f=f_{-\mu}+\mu\phi$. Although f may not have bounded L_f , the shifted function $f_{-\mu}$ does. Equivalently, the non-Lipschitz component of f can serve as the mirror function, with the remainder being smooth. See the **Entropic Mirror Descent** example in the numerical section.

Another important class is when the mirror function ϕ has continuous Hessian in the ball $B(x^*, R)$, which covers the **Quartic Objective** example in the numerical section.

Theorem 3.6. Assume f satisfies (A1), and $\phi \in C^{2,1}(V)$ with ϕ being 1-strongly convex. When $(x, \hat{x}, y, \hat{y}) \in B(x^*, R)$, (A2) holds with

$$C_{f,\phi} \le (L - \mu) \left(1 + 2L_{\nabla^2 \phi} R \right),$$

where $L_{\nabla^2 \phi}$ is the Lipschitz constant of $\nabla^2 \phi$ restricted to $B(x^*, R)$.

Proof. Take $A = \nabla^2 \phi(\xi_y)$ for some point ξ_y between y and \hat{y} such that $D_\phi(y,\hat{y}) = \frac{1}{2} \|y - \hat{y}\|_A^2$. As ϕ is 1-strongly convex, $\|e\| \le \|e\|_A$ for any $e \in V$. Let ξ_x be a point between x and \hat{x} satisfying $\langle \nabla \phi(x) - \nabla \phi(\hat{x}), x - \hat{x} \rangle = \|x - \hat{x}\|_{\nabla^2 \phi(\xi_x)}^2$. By assumption $\xi_x, \xi_y \in B(x^*, R)$ and

$$|\|e\|_{\nabla^2\phi(\xi_x)}^2 - \|e\|_{\nabla^2\phi(\xi_y)}^2| \le L_{\nabla^2\phi}\|\xi_x - \xi_y\|\|e\|^2 \le 2L_{\nabla^2\phi}R\|e\|^2 \quad \forall e \in B(x^*, R).$$

We use the relative smoothness to get

$$\begin{aligned} |\langle \nabla f_{-\mu}(x) - \nabla f_{-\mu}(\hat{x}), \ x - \hat{x} \rangle| &\leq (L - \mu) \|x - \hat{x}\|_{\nabla^2 \phi(\xi_x)}^2 \\ &\leq (L - \mu) \left(\|x - \hat{x}\|_{\nabla^2 \phi(\xi_y)}^2 + L_{\nabla^2 \phi} \|\xi_x - \xi_y\| \|x - \hat{x}\|^2 \right) \\ &\leq (L - \mu) (1 + 2L_{\nabla^2 \phi} R) \|x - \hat{x}\|_A^2, \end{aligned}$$

which implies $L_{f_{-\mu}}(A) \leq (L-\mu)(1+2L_{\nabla^2\phi}R)$. Applying Theorem 3.5 to get the desired estimate. \Box

In particular, when ϕ is quadratic, $\nabla^2 \phi$ is constant and $L_{\nabla^2 \phi} = 0$, recovering a sharper estimate $C_{f,\phi} \leq L - \mu$. In our implementation, we take $C_{f,\phi} = L$ as an effective estimate. Again adaptivity of $C_{f,\phi}$ may further improve the convergence.

Extension to convex optimization For convex objectives, i.e., $\mu = 0$, we consider the following perturbed flow:

$$\begin{cases} x' = y - x, \\ (\nabla \phi(y))' = \epsilon^{-1} \left[\epsilon \left(\nabla \phi(x) - \nabla \phi(y) \right) - \nabla f(x) \right], \end{cases}$$
 (23)

for a fixed perturbation level $\epsilon > 0$. In dual variables, this becomes

$$x' = y - x$$
, $\eta' = -\epsilon^{-1} \nabla f(x) + \chi - \eta$,

where $\chi = \nabla \phi(x)$ and $\eta = \nabla \phi(y)$. Owing to the variable splitting structure, the perturbation does not change the stationary points of the original system.

To accelerate convergence, we employ a homotopy strategy that gradually decreases the perturbation level. Starting from ϵ_0 , we run accelerated mirror descent for m_k iterations at level ϵ_k , and update

$$\epsilon_{k+1} = \frac{\epsilon_k}{2}, \qquad m_{k+1} = \sqrt{2} \, m_k.$$

This geometric schedule ensures an accelerated sublinear convergence rate with complexity $\mathcal{O}(C_{f,\phi}/k^2)$. The convergence analysis follows (Chen et al., 2025) and is detailed in Appendix A.

Extension to composite optimization Consider the composite optimization problem

$$\min_{x \in \mathbb{R}^n} \quad F(x) := f(x) + g(x), \tag{24}$$

where f satisfies the relative smoothness and convexity condition (10), and g is convex but may be non-smooth, with a well-defined generalized proximal operator. The mirror function ϕ is chosen for f.

We extend the accelerated mirror descent method to the composite setting by discretizing the modified flow (2). The only change lies in Line 4 of Algorithm 1, where y_{k+1} is computed via a generalized proximal operator:

$$y_{k+1} = \arg\min_{y \in \mathbb{R}^n} (1 + \alpha) \phi(y) + \frac{\alpha}{\mu} g(y) - \left\langle -\frac{\alpha}{\mu} \nabla f(x_k) + \alpha \nabla \phi(x_k) + \nabla \phi(y_k), y \right\rangle. \tag{25}$$

Since y is treated implicitly, the convergence analysis remains the same as in the smooth case, showing the flexibility of the VOS framework. For details, see Appendix B. A perturbation and homotopy argument can be further applied if $\mu = 0$.

4 NUMERICAL EXAMPLES

We evaluate the performance of Acc-MD on a variety of convex optimization problems. All experiments were conducted in MATLAB R2023a on a desktop with an Intel Core i5-7200U CPU (2.50 GHz) and 8 GB RAM. Random seeds were fixed for reproducibility.

We compare Acc-MD against several state-of-the-art first-order methods from the literature. In all tested scenarios, Acc-MD consistently outperforms competing algorithms by a large margin.

Entropic Mirror Descent This example arises in log-linear models in supervised machine learning. According to Collins et al. (2008), the dual function of the log-linear objective can be written as

$$f(x) = \sum_{i=1}^{d} x_i \log x_i + \frac{1}{2} x^{\top} A x,$$

where $x = (x_i) \in \mathbb{R}^d$ denotes the probability of labels, and $A = \mathbf{g}\mathbf{g}^{\top}$ with a given $\mathbf{g} \in \mathbb{R}^d$.

We take Shannon's entropy function $\phi(x) = \sum_i x_i \log x_i$ as the mirror function. It is well known that the Bregman divergence $D_{\phi}(x,z)$ can be reformulated as the KL-divergence between two discrete probability measures:

$$D_{\phi}(x, z) = \text{KL}(x, z) = \sum_{i} x_{i} \log \left(\frac{x_{i}}{z_{i}}\right).$$

(Collins et al., 2008, Lemma 7) proved that f is 1-relatively strongly convex and $(1+|A|_{\infty})$ -relatively smooth with respect to ϕ , where $|A|_{\infty}$ refers to the largest entry in magnitude across all rows and columns of A. Thus, assumption (A1) holds.

To verify Assumption (A2), we note that $f_{-\mu} = \frac{1}{2}x^{\top}Ax$ is quadratic and $L_{f_{-\mu}}(A) = 1$. By Pinsker's inequality, $\|y - \hat{y}\|_1 \le \sqrt{2 \operatorname{KL}(\hat{y}, y)}$, we obtain

$$\|y - \hat{y}\|_{A} \le \|\mathbf{g}\|_{2} \|y - \hat{y}\|_{2} \le \|\mathbf{g}\|_{2} \|y - \hat{y}\|_{1} \le \|\mathbf{g}\|_{2} \sqrt{2 \operatorname{KL}(\hat{y}, y)} = \|\mathbf{g}\|_{2} \sqrt{2 D_{\phi}(\hat{y}, y)}.$$

Therefore, Assumption (A2) holds with constant

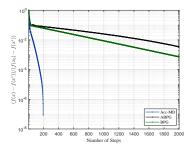
$$C_{f,\phi} = \frac{L_{f-\mu}}{\mu_{\phi}} = \|\mathbf{g}\|_2^2 = \|A\|_2.$$

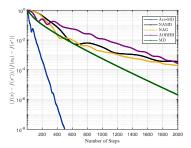
By Theorem 3.4, Acc-MD exhibits accelerated linear convergence rate

$$(1 + \sqrt{\|A\|_2})^{-1} = (1 + \|\mathbf{g}\|_2)^{-1},$$

when applied to this example, while other existing entropic mirror descent methods achieve at most the sublinear rate $\mathcal{O}(1/k^2)$.

We compare Acc-MD with Accelerated Bregman Proximal Gradient (ABPG) and Bregman Proximal Gradient (BPG) from Hanzely et al. (2021) (Figure 1a). For this example, the TSE component is $\gamma=1$, and ABPG achieves only a non-accelerated rate of $\mathcal{O}(1/k)$, converging slightly slower than BPG. This agrees with the results in (Hanzely et al., 2021, Section 6.1.1). Acc-MD significantly outperforms the other methods and shows a steep and stable error curve.





- (a) Log error curves on the entropic MD problem.
- (b) Log error curves on the quartic problem.

Figure 1: Comparison of log error curves for different problems.

Quartic Objective We evaluate Acc-MD on smooth convex minimization problems and compare it with the following first-order methods:

- 1. Nesterov's accelerated mirror descent (NAMD) (Nesterov, 2005);
- 2. Nesterov's accelerated gradient (NAG) (Nesterov, 1983) with step size 1/(k+3);
- 3. Accelerated over-relaxation heavy ball (AOR-HB) (Wei and Chen, 2025);
- 4. Mirror descent (MD) (Lu et al., 2018).

All methods use the stopping criterion $\|\nabla f(x_k)\| \le \text{tol} \cdot \|\nabla f(x_0)\|$ with tolerance $\text{tol} = 10^{-6}$.

We consider the smooth, strongly convex objective in Section 2.1 of (Lu et al., 2018):

$$f(x) = \frac{1}{4} ||Ex||_{2}^{4} + \frac{1}{4} ||Ax - b||_{4}^{4} + \frac{1}{2} ||Cx - d||_{2}^{2},$$

where C and E are $n \times n$ positive definite matrices. The mirror function is chosen as

$$\phi(x) = \frac{1}{4} ||x||_2^4 + \frac{1}{2} ||x||_2^2.$$

The update step takes the form $y_{k+1} = \nabla \phi^*(c) = \theta c$, where $\theta > 0$ solves the cubic equation $\|c\|_2^2 \theta^3 + \theta - 1 = 0$, and can be calculated by root function in MATLAB, and $c = \frac{1}{1+\alpha} \left[\alpha \nabla \phi(x_k) + \nabla \phi(y_k) - \frac{\alpha}{\mu} \nabla f(x_k) \right]$. We refer to (Lu et al., 2018) for the full derivation.

Let λ_C and λ_E denote the smallest eigenvalues of C and E, respectively. Since $\|\nabla^2 f\|$ grows quadratically with $\|x\|_2$, the gradient ∇f is not globally Lipschitz. To apply NAG and AOR-HB, we assume $\|x\|_2 \leq R$ and estimate the global parameters as

$$L_f = (3\|E\|^4 + 3\|A\|^4)R^2 + 6\|A\|^3\|b\|_2R + 3\|A\|^2\|b\|_2^2 + \|C\|^2, \qquad \mu = \lambda_C^2.$$

According to Lu et al. (2018), (A1) holds with the relative smoothness constant $L=L_f$ when R=1, and the relative strong convexity constant $\mu=\min\left\{\lambda_E^4/3,\ \lambda_C^2\right\}$.

Since ϕ is smooth, we can apply Theorem 3.6 to verify (A2).

Set n=256, $A\sim \frac{1}{\sqrt{n}}\mathcal{N}(\mathbf{0},I_{n\times n})$, $C=I_n+C_0C_0^\top/n$ with $C_0\sim \mathcal{N}(\mathbf{0},I_{n\times n})$, $E=2I_n+E_0E_0^\top/n$ with $E_0\sim \mathcal{N}(\mathbf{0},I_{n\times n})$, b=0, and $d\sim \mathrm{Unif}(0,1)^n$. This gives $\mu=\mu_\phi=1$, L remains fixed, while L_f grows with $R=\|d\|_2$, highlighting the benefit of using relative smoothness. This is experimentally verified in Figure 1b, where Acc-MD and MD converge faster and more stably than all other non-mirror-type methods. Moreover, Acc-MD outperforms MD, making it the best method for this problem.

We provide more examples, including applications to composite convex optimization (LASSO) and constrained convex optimization (quadratic optimization on the simplex), in Appendix B.

REFERENCES

- Heinz H Bauschke, Jérôme Bolte, and Marc Teboulle. A descent lemma beyond Lipschitz gradient continuity: first-order methods revisited and applications. *Mathematics of Operations Research*, 42(2):330–348, 2017.
- Amir Beck and Marc Teboulle. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31(3):167–175, 2003. ISSN 0167-6377. doi: https://doi.org/10.1016/S0167-6377(02)00231-6.
- Benjamin Birnbaum, Nikhil R. Devanur, and Lin Xiao. Distributed algorithms via gradient descent for Fisher markets. In *Proceedings of the 12th ACM Conference on Electronic Commerce*, EC '11, page 127–136, New York, NY, USA, 2011. Association for Computing Machinery. ISBN 9781450302616. doi: https://doi.org/10.1145/1993574.1993594.
- Sébastien Bubeck. Convex optimization: Algorithms and complexity. *Foundations and Trends® in Machine Learning*, 8(3–4):231–357, November 2015. ISSN 1935-8237. doi: https://doi.org/10.1561/2200000050.
- Gong Chen and Marc Teboulle. Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM Journal on Optimization*, 3(3):538–543, 1993. doi: https://doi.org/10.1137/0803026.
- Long Chen and Hao Luo. A unified convergence analysis of first order convex optimization methods via strong Lyapunov functions, 2021. URL https://arxiv.org/abs/2108.00132.
- Long Chen, Luo Hao, and Jingrong Wei. Accelerated gradient methods through variable and operator splitting, 2025. URL https://arxiv.org/abs/2505.04065.
- Michael Collins, Amir Globerson, Terry Koo, Xavier Carreras, and Peter L. Bartlett. Exponentiated gradient algorithms for conditional random fields and max-margin Markov networks. *Journal of Machine Learning Research*, 9(58):1775–1822, 2008. URL http://jmlr.org/papers/v9/collins08a.html.
- Radu-Alexandru Dragomir, Adrien B. Taylor, Alexandre d'Aspremont, and Jérôme Bolte. Optimal complexity and certification of Bregman first-order methods. *Mathematical Programming*, 194(1): 41–83, 2022. doi: https://doi.org/10.1007/s10107-021-01618-1.
- Filip Hanzely, Peter Richtárik, and Lin Xiao. Accelerated Bregman proximal gradient methods for relatively smooth convex optimization. *Computational Optimization and Applications*, 79(2): 405–440, June 2021. doi: https://doi.org/10.1007/s10589-021-00273-.
- Walid Krichene, Alexandre Bayen, and Peter L Bartlett. Accelerated mirror descent in continuous and discrete time. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL https://proceedings.neurips.cc/paper_files/paper/2015/file/f60bb6bb4c96d4df93c51bd69dcc15a0-Paper.pdf.
- Haihao Lu, Robert M. Freund, and Yurii Nesterov. Relatively smooth convex optimization by first-order methods, and applications. *SIAM Journal on Optimization*, 28(1):333–354, 2018. doi: https://doi.org/10.1137/16M1099546.
- Arkadij Semenovič Nemirovskij and David Borisovich Yudin. *Problem Complexity and Method Efficiency in Optimization*. A Wiley-Interscience publication. Wiley, 1983. ISBN 9780471103455.
- Yurii Nesterov. A method of solving a convex programming problem with convergence rate $o(\frac{1}{k^2})$. *Doklady Akademii Nauk SSSR*, 269(3):543–547, 1983.
- Yurii Nesterov. Smooth minimization of non-smooth functions. *Mathematical Programming*, 103(1): 127–152, 2005. doi: https://doi.org/10.1007/s10107-004-0552-5.
- Stanley Osher, Feng Ruan, Jiechao Xiong, Yuan Yao, and Wotao Yin. Sparse recovery via differential inclusions. *Applied and Computational Harmonic Analysis*, 41(2):436–469, 2016. ISSN 1063-5203. doi: https://doi.org/10.1016/j.acha.2016.01.002. Sparse Representations with Applications in Imaging Science, Data Analysis, and Beyond, Part II.

Weijie Su, Stephen Boyd, and Emmanuel J. Candès. A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights. Journal of Machine Learning Research, 17 (153):1-43, 2016. URL http://jmlr.org/papers/v17/15-084.html. Marc Teboulle. A simplified view of first order methods for optimization. *Mathematical Programming*, 170(1):67–96, 2018. doi: https://doi.org/10.1007/s10107-018-1284-2. Jingrong Wei and Long Chen. Accelerated over-relaxation Heavy-Ball method: Achieving global accelerated convergence with broad generalization. In The Thirteenth International Confer-ence on Learning Representations, 2025. URL https://openreview.net/forum?id= SWEqzy7IQB. Andre Wibisono, Ashia C. Wilson, and Michael I. Jordan. A variational perspective on accelerated methods in optimization. Proceedings of the National Academy of Sciences, 113(47):E7351–E7358, 2016. doi: https://doi.org/10.1073/pnas.1614734113. Ya xiang Yuan and Yi Zhang. Analysis accelerated mirror descent via high-resolution ODEs, 2023. URL https://arxiv.org/abs/2308.03242. Yi Zhou, Yingbin Liang, and Lixin Shen. A simple convergence analysis of Bregman proximal gradient algorithm. Computational Optimization and Applications, 73(3):903-912, 2019. doi: https://doi.org/10.1007/s10589-019-00092-y.

APPENDIX A: EXTENSION TO CONVEX OPTIMIZATION

For convex functions, i.e, $\mu = 0$, we consider the perturbed system

$$\begin{cases} x' = y - x, \\ (\nabla \phi(y))' = \epsilon^{-1} \left[\epsilon (\nabla \phi(x) - \nabla \phi(y)) - \nabla f(x) \right], \end{cases}$$
 (26)

for some fixed $\epsilon > 0$. With the dual variables, the perturbed flow is equivalent to

$$x' = y - x$$
, $\eta' = -\epsilon^{-1} \nabla f(x) + \chi - \eta$.

Consider the Lyapunov function

$$\mathcal{E}(x,\eta;\epsilon) := D_f(x,x^*) + \epsilon D_{\phi^*}(\eta,\chi^*). \tag{27}$$

We treat ϵ as a fixed parameter and take derivative with respect to x and η .

Lemma 4.1. Let \mathcal{E} be the Lyapunov function (27), and let \mathcal{G} be the vector field of the perturbed acclerated mirror descent flow (26). The following perturbed strong Lyapunov property holds:

$$-\nabla \mathcal{E}(x,\eta;\epsilon) \cdot \mathcal{G}(x,\eta) = \mathcal{E}(x,\eta;\epsilon) + D_f(x^*,x) + \epsilon D_{\phi^*}(\chi,\eta) - \epsilon D_{\phi^*}(\chi,\chi^*). \tag{28}$$

Proof. By direct calculations,

$$-\nabla \mathcal{E}(x,\eta;\epsilon) \cdot \mathcal{G}(x,\eta) = \langle \nabla f(x) - \nabla f(x^*), x - x^* \rangle + \epsilon (\nabla \phi^*(\eta) - \nabla \phi^*(\chi^*), \eta - \chi).$$

Using the three-point identity of the Bregaman divergence, the additional cross term expands as:

$$\epsilon \langle \nabla \phi^*(\eta) - \nabla \phi^*(\chi^*), \eta - \chi \rangle = \epsilon \left(D_{\phi^*}(\eta, \chi^*) + D_{\phi^*}(\chi, \eta) - D_{\phi^*}(\chi, \chi^*) \right).$$

Therefore,

$$-\nabla \mathcal{E}(x,\eta;\epsilon) \cdot \mathcal{G}(x,\eta) = \mathcal{E}(x,y;\epsilon) + D_f(x^*,x) + \epsilon D_{\phi^*}(\chi,\eta) - \epsilon D_{\phi^*}(\chi,\chi^*).$$

Consider the perturbed AOR accelerated mirror descent method:

$$\frac{x_{k+1} - x_k}{\alpha} = y_k - x_{k+1},\tag{29a}$$

$$\frac{\nabla \phi(y_{k+1}) - \nabla \phi(y_k)}{\alpha} = -\epsilon^{-1} (2\nabla f(x_{k+1}) - \nabla f(x_k)) + \nabla \phi(x_{k+1}) - \nabla \phi(y_{k+1}). \tag{29b}$$

An equivalent but computation favorable form is given as

$$x_{k+1} = \frac{1}{1+\alpha} (x_k + \alpha y_k),$$

$$y_{k+1} = \arg\min_{y \in \mathbb{R}^n} (1+\alpha) \phi(y) - \left\langle \alpha \nabla \phi(x_{k+1}) + \nabla \phi(y_k) - \frac{\alpha}{\epsilon} (2\nabla f(x_{k+1}) - \nabla f(x_k)), y \right\rangle.$$
(30)

Introduce the dual variables $\eta_k = \nabla \phi(y_k), \chi_k = \nabla \phi(x_k), k = 0, 1, 2, \cdots$ to rewrite (29b) as

$$\frac{\eta_{k+1} - \eta_k}{\alpha} = -\frac{1}{\epsilon} (2\nabla f(x_{k+1}) - \nabla f(x_k)) + \chi_{k+1} - \eta_{k+1}. \tag{31}$$

For four points x, \hat{x}, y, \hat{y} , introduce the term

$$\mathcal{B}^{\alpha}(x,\hat{x},\hat{y},y;\epsilon) = D_f(x,\hat{x}) + \epsilon D_{\phi}(\hat{y},y) + \alpha \left\langle \nabla f(x) - \nabla f(\hat{x}), y - \hat{y} \right\rangle. \tag{32}$$

Define the modified Lyapunov function

$$\mathcal{E}^{\alpha}(x, y; \epsilon) := \mathcal{B}^{\alpha}(x, x^*, x^*, y; \epsilon) := \mathcal{E}(x, \eta; \epsilon) + \alpha \left\langle \nabla f(x) - \nabla f(x^*), y - x^* \right\rangle. \tag{33}$$

For simplicity, we shall use $\mathcal{B}^{\alpha}(x,\hat{x},\hat{y},y)$ and $\mathcal{E}^{\alpha}(x,y)$ when there is no confusion to skip ϵ .

Lemma 4.2. Let (x_k, y_k) be the sequence generated by perturbed Acc-MD (29), then it holds

$$\mathcal{E}^{\alpha}(x_{k+1}, y_{k+1}) - \mathcal{E}^{\alpha}(x_k, y_k) \leq -\alpha \mathcal{E}^{\alpha}(x_{k+1}, y_{k+1}) - \alpha D_f(x^*, x_{k+1}) + \alpha \left\langle \nabla f(x_{k+1}) - \nabla f(x^*), y_{k+1} - x^* \right\rangle - \mathcal{B}^{\alpha}(x_k, x_{k+1}, y_{k+1}, y_k) + \alpha \epsilon D_{\phi}(x^*, x_{k+1})$$
(34)

Proof. Denote $z = (x, \eta)$. Expand the difference $\mathcal{E}(x, \eta)$ at (x_{k+1}, η_{k+1}) ,

$$\mathcal{E}(x_{k+1}, \eta_{k+1}) - \mathcal{E}(x_k, \eta_k) = \langle \nabla \mathcal{E}(\boldsymbol{z}_{k+1}), \boldsymbol{z}_{k+1} - \boldsymbol{z}_k \rangle - D_{\mathcal{E}}(\boldsymbol{z}_k, \boldsymbol{z}_{k+1})$$

$$= \alpha \langle \nabla \mathcal{E}(\boldsymbol{z}_{k+1}; \epsilon), \mathcal{G}(z_{k+1}) \rangle - D_{\mathcal{E}}(z_k, z_{k+1})$$

$$- \alpha \langle \nabla f(x_{k+1}) - \nabla f(x^*), y_{k+1} - y_k \rangle$$

$$- \alpha \langle y_{k+1} - x^*, \nabla f(x_{k+1}) - \nabla f(x_k) \rangle$$

We write the cross term as

$$\langle \nabla f(x_{k+1}) - \nabla f(x^*), y_{k+1} - y_k \rangle + \langle y_{k+1} - x^*, \nabla f(x_{k+1}) - \nabla f(x_k) \rangle$$

$$= \langle \nabla f(x_{k+1}) - \nabla f(x^*), y_{k+1} - x^* \rangle - \langle \nabla f(x_k) - \nabla f(x^*), y_k - x^* \rangle$$

$$+ \langle \nabla f(x_{k+1}) - \nabla f(x_k), y_{k+1} - y_k \rangle$$

Then use identity (28) to expand $\langle \nabla \mathcal{E}(z_{k+1}), \mathcal{G}(z_{k+1}) \rangle$ we get

$$\mathcal{E}^{\alpha}(x_{k+1}, y_{k+1}) - \mathcal{E}^{\alpha}(x_k, y_k) = -\alpha \mathcal{E}(x_{k+1}, \eta_{k+1}) - \alpha D_f(x^*, x_{k+1}) - \alpha \epsilon D_{\phi}(y_{k+1}, x_{k+1}) + \alpha \epsilon D_{\phi}(x^*, x_{k+1}) - D_{\mathcal{E}}(z_k, z_{k+1}) - \alpha \left\langle \nabla f(x_{k+1}) - \nabla f(x_k), y_{k+1} - y_k \right\rangle = -\alpha \mathcal{E}^{\alpha}(x_{k+1}, y_{k+1}) - \alpha D_f(x^*, x_{k+1}) - \alpha \epsilon D_{\phi}(y_{k+1}, x_{k+1}) + \alpha \epsilon D_{\phi}(x^*, x_{k+1}) + \alpha \left\langle \nabla f(x_{k+1}) - \nabla f(x^*), y_{k+1} - x^* \right\rangle - \mathcal{B}^{\alpha}(x_k, x_{k+1}, y_k, y_{k+1}),$$

where noted that $D_{\phi^*}(\chi_{k+1}, \chi^*) = D_{\phi}(x^*, x_{k+1})$. Drop the negative terms and rearrangement we get the desired result.

We make an analogy to the assumption (A2).

Assumption 4.3. There exists $C_{f,\phi} > 0$ such that the following relative Cauchy Schwarz inequality holds

$$|\langle \nabla f(x) - \nabla f(\hat{x}), y - \hat{y} \rangle| \le 2\sqrt{C_{f,\phi}} D_f^{1/2}(x, \hat{x}) D_{\phi}^{1/2}(\hat{y}, y), \quad \forall x, \hat{x}, y, \hat{y} \in V.$$

Lemma 4.4. Let Assumption 4.3 hold. Then for $|\alpha| \leq \sqrt{\epsilon/C_{f,\phi}}$, we have $\mathcal{B}^{\alpha}(x,\hat{x},\hat{y},y) \geq 0$.

Proof. From the Cauchy Schwarz inequality, we can derive

$$|\langle \nabla f(x) - \nabla f(\hat{x}), y - \hat{y} \rangle| \le \sqrt{\frac{C_{f,\phi}}{\epsilon}} \left[D_f(x, \hat{x}) + \epsilon D_\phi(\hat{y}, y) \right]. \tag{35}$$

 \Box

Theorem 4.5 (Convergence of perturbed Acc-MD method). Suppose f is convex and relative Cauchy Schwarz inequality holds with constant $C_{f,\phi}$. For any $\epsilon > 0$, let (x_k, y_k) be generated by scheme (29) with initial value (x_0, y_0) , $\eta_k = \nabla \phi(y_k)$, and step size $0 < \alpha = \sqrt{\epsilon/C_{f,\phi}}$. Assume there exists R > 0 such that

$$D_{\phi}(x^*, x_k) \le R/2, \quad D_{\phi}(x^*, y_k) \le R/2, \quad \forall k \ge 0.$$
 (36)

Then we have the linear convergence with perturbation

$$\mathcal{E}^{\alpha}(x_k, y_k, \epsilon) \le \left(\frac{1}{1 + \sqrt{\epsilon/C_{f,\phi}}}\right)^k \mathcal{E}^{\alpha}(x_0, y_0, \epsilon) + \epsilon R, \quad k \ge 0.$$
 (37)

Proof. By Lemma 4.4,

$$\mathcal{E}^{\alpha}(x_{k+1}, y_{k+1}) - \mathcal{E}^{\alpha}(x_k, y_k) \leq -\alpha \mathcal{E}^{\alpha}(x_{k+1}, y_{k+1}) - \alpha D_f(x^*, x_{k+1}) - \alpha \epsilon D_{\phi}(x^*, y_{k+1}) + \alpha \left\langle \nabla f(x_{k+1}) - \nabla f(x^*), y_{k+1} - x^* \right\rangle - \mathcal{B}^{\alpha}(x_k, x_{k+1}, y_{k+1}, y_k) + \alpha \epsilon (D_{\phi}(x^*, x_{k+1}) + \epsilon D_{\phi}(x^*, y_{k+1})).$$
(38)

For $\alpha = \sqrt{\epsilon/C_{f,\phi}}$, we can drop negative terms from the inequality (38) to get the accelerated linear convergence

$$\begin{split} \mathcal{E}^{\alpha}(x_{k+1}, y_{k+1}) &\leq \frac{1}{1+\alpha} \mathcal{E}^{\alpha}(x_k, y_k) + \frac{\alpha \epsilon}{1+\alpha} (D_{\phi}(x^*, x_{k+1}) + D_{\phi}(x^*, y_{k+1})) \\ &\leq \frac{1}{1+\alpha} \mathcal{E}^{\alpha}(x_k, y_k, \epsilon) + \frac{\alpha \epsilon}{1+\alpha} R \\ &\leq \left(\frac{1}{1+\alpha}\right)^{k+1} \mathcal{E}^{\alpha}(x_0, y_0, \epsilon) + \frac{\alpha \epsilon R}{1+\alpha} \sum_{i=0}^{k} \frac{1}{(1+\alpha)^i}. \end{split}$$

Summing up the geometric series we get the desired result.

To achieve an accuracy $\mathcal{E}^{\alpha}(x_k, y_k) = O(\epsilon)$, the number of iterations is bounded by

$$\left(1+\sqrt{\epsilon/C_{f,\phi}}\right)^{-k}=O(\epsilon)\quad\Longrightarrow\quad k=O\left(\sqrt{\frac{C_{f,\phi}}{\epsilon}}|\ln\epsilon|\right).$$

Compared to the dominant complexity $O(\epsilon^{-1/2})$, the logarithmic factor $O(|\ln \epsilon|)$ is negligible. This establishes the nearly optimal complexity of accelerated gradient methods.

Since the perturbation will not change the equilibrium point x^* , we can choose strictly decreasing ϵ_{k+1} and use the homotopy argument to remove the $|\ln \epsilon|$ dependence (Chen et al., 2025, Theorem 8.4). The modified algorithm is summarized in Algorithm ??. Consequently, the method achieves an effective sublinear convergence rate of $\mathcal{O}(1/k^2)$ in terms of gradient evaluations.

Algorithm 2 AccMD with homotopy perturbation.

- 1: Parameters: Inititial value and tolerance (x_0,y_0,ϵ_0) and termination tolerance ϵ .
- 2: Set k = 0 and $m_0 = (\sqrt{C_{f,\phi}} + \sqrt{\epsilon_0}) \ln(2(R+1)) \epsilon_0^{-1/2}$
- 3: while $\epsilon_k > \epsilon$ do
- 4: $\epsilon_{k+1} = \epsilon_k/2$, $m_{k+1} = \sqrt{2} m_k$
- 5: Apply perturbed AccMD scheme (30) with the initial value (x_k, y_k) , the parameter ϵ_{k+1} and the step size $\alpha = \sqrt{\epsilon_{k+1}/C_{f,\phi}}$ for m_{k+1} iterations to get (x_{k+1}, y_{k+1})
- 6: k = k + 1
- 7: end while
- 8: **return** (x_k, y_k)

APPENDIX B: EXTENSION TO COMPOSITE OPTIMIZATION

Consider the composite optimization problem

$$\min_{x \in \mathbb{R}^n} \quad F(x) := f(x) + g(x), \tag{39}$$

where f satisfies the relative smoothness and convexity condition, and g is convex but possibly non-smooth, with a well-defined generalized proximal operator. We propose the accelerated mirror descent flow for composite case

$$\begin{cases} x' = y - x, \\ -\mu \left(\nabla \phi(y)\right)' + \mu \left(\nabla \phi(x) - \nabla \phi(y)\right) - \nabla f(x) \in \partial g(y), \end{cases}$$
(40)

with initial conditions $x(0) = x_0$, $y(0) = y_0$. Introducing dual variables $\chi = \nabla \phi(x)$ and $\eta = \nabla \phi(y)$, the flow (40) becomes

$$x' = y - x$$
, $\eta' = -\mu_{\phi}^{-1} [\nabla f_{-\mu}(x) + q(y)] - \eta$,

where $f_{-\mu} := f - \mu_{\phi} \phi$ and $q(y) \in \partial g(y)$ is one sub-gradient of g at y.

We make an analysis analogous to the smooth convex case. Define a Lyapunov function:

$$\mathcal{E}(x,\eta) := D_{f_{-\mu}}(x,x^*) + \mu D_{\phi^*}(\eta,\chi^*), \tag{41}$$

where x^* is a minimizer of f + g and $\chi^* = \nabla \phi(x^*)$.

Lemma 4.6. Let $\mathcal{E}(x,\eta)$ be defined by (41), and define the vector field $\mathcal{G}(x,\eta) = (y - x, -\mu_{\phi}^{-1}[\nabla f_{-\mu}(x) + q(y)] - \eta)$, where $q(y) \in \partial g(y)$, and the dual variables satisfy $\chi = \nabla \phi(x)$ and $\eta = \nabla \phi(y)$. Then

$$-\nabla \mathcal{E}(x,\eta) \cdot \mathcal{G}(x,\eta) \ge \mathcal{E}(x,\eta) + D_{f_{-\mu}}(x^*,x) + \mu_{\phi} D_{\phi}(y,x^*). \tag{42}$$

As a consequence, any solution (x(t), y(t)) of the flow (40) satisfies the exponential decay bound

$$\mathcal{E}(x(t), \eta(t)) \le e^{-t} \mathcal{E}(x(0), \eta(0)). \tag{43}$$

Proof. First, observe that at minimum x^* , $0 \in \nabla f(x^*) + \partial g(x^*)$, so there is $q(x^*) \in \partial g(x^*)$ that $\nabla f(x^*) + q(x^*) = 0$ holds. From the gradients of the Lyapunov function:

$$\partial_x \mathcal{E} = \nabla f_{-\mu}(x) - \nabla f_{-\mu}(x^*), \quad \partial_\eta \mathcal{E} = \mu_\phi \left(\nabla \phi^*(\eta) - \nabla \phi^*(\chi^*) \right) = \mu_\phi(y - x^*),$$

we compute

$$\begin{split} &-\nabla \mathcal{E}(x,\eta) \cdot \mathcal{G}(x,\eta) \\ &= \left\langle \nabla f_{-\mu}(x) - \nabla f_{-\mu}(x^*), x - y \right\rangle + \mu_{\phi} \left\langle y - x^*, \, \mu_{\phi}^{-1}(\nabla f_{-\mu}(x) + q(y)) + \eta \right\rangle \\ &= \left\langle \nabla f_{-\mu}(x) - \nabla f_{-\mu}(x^*), x - x^* \right\rangle + \mu_{\phi} \left\langle \nabla \phi^*(\eta) - \nabla \phi^*(\chi^*), \eta - \chi^* \right\rangle \\ &= \left\langle \nabla f_{-\mu}(x) - \nabla f_{-\mu}(x^*), x - x^* \right\rangle + \left\langle y - x^*, \nabla f(x^*) + q(y) \right\rangle + \mu_{\phi} \left\langle \nabla \phi^*(\eta) - \nabla \phi^*(\chi^*), \eta - \chi^* \right\rangle \\ &= D_{f_{-\mu}}(x, x^*) + D_{f_{-\mu}}(x^*, x) + \left\langle y - x^*, q(y) - q(x^*) \right\rangle + \mu_{\phi} D_{\phi^*}(\chi^*, \eta) + \mu_{\phi} D_{\phi^*}(\eta, \chi^*) \\ &\geq \mathcal{E}(x, \eta) + D_{f_{-\mu}}(x^*, x) + \mu_{\phi} D_{\phi}(y, x^*), \end{split}$$

where the last inequality follows from convexity of g as $\langle y - x^*, q(y) - q(x^*) \rangle \geq 0$.

Since $D_{f_{-\mu}}(x^*,x) \ge 0$ and $D_{\phi}(y,x^*) \ge 0$ by convexity of $f_{-\mu}$ and ϕ , we obtain the inequality

$$\nabla \mathcal{E}(x,\eta) \cdot \mathcal{G}(x,\eta) < -\mathcal{E}(x,\eta),$$

from which the exponential decay (43) follows by Grönwall's inequality.

We generalize the Acc-MD scheme for composite case

$$\frac{x_{k+1} - x_k}{\alpha} = 2y_{k+1} - y_k - x_{k+1},\tag{44a}$$

$$\frac{\nabla \phi(y_{k+1}) - \nabla \phi(y_k)}{\alpha} \in -\frac{1}{\mu} [\nabla f(x_k) + \partial g(y_{k+1})] + \nabla \phi(x_k) - \nabla \phi(y_{k+1}). \tag{44b}$$

An equivalent computation favorable form is given as Algorithm 3.

As the subgradient of g is evaluated implicitly in scheme (44), the convergence analysis becomes identical to the smooth convex case. Combining with the convergence result of ODE flow Lemma 4.6, we derive the following theorem for Algorithm 3:

Theorem 4.7 (Convergence of Acc-MD method for composite optimization). Suppose f is μ -relatively convex and relative Cauchy Schwarz inequality holds with constant $C_{f,\phi}$ and g is convex. Let (x_k, y_k) be generated by scheme (44) with initial value (x_0, y_0) , $\eta_k = \nabla \phi(y_k)$, and step size $\alpha = \sqrt{\mu/C_{f,\phi}}$. Then there exists a constant $C_0 = C_0(x_0, y_0, \mu, C_{f,\phi})$ so that we have the accelerated linear convergence

$$D_{f_{\mu}}(x_{k+1}, x^*) + \mu D_{\phi^*}(\eta_{k+1}, \chi^*) \le C_0 \left(\frac{1}{1 + \sqrt{\mu/C_{f,\phi}}}\right)^k, \quad k \ge 1.$$

Algorithm 3 Accelerated mirror descent (Acc-MD) method for composite optimization

1: **Parameters:** $x_0, y_0 \in \mathbb{R}^n, \mu, C_{f,\phi}$.

2: Set $\alpha = \sqrt{\mu/C_{f,\phi}}$.

3: **for** $k = 0, 1, 2, \dots$ **do**

4: $y_{k+1} = \arg\min_{y \in \mathbb{R}^n} (1+\alpha) \, \phi(y) + \frac{\alpha}{\mu} g(y) - \left\langle \alpha \nabla \phi(x_k) + \nabla \phi(y_k) - \frac{\alpha}{\mu_\phi} \nabla f(x_k), \, y \right\rangle.$

5: $x_{k+1} = \frac{1}{1+\alpha} [x_k + \alpha(2y_{k+1} - y_k)].$

6: end for

LASSO problem We consider the over-parameterized LASSO problem:

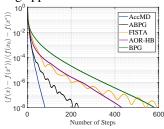
$$\min_{x \in \mathbb{R}^d} F(x) := \frac{1}{2} ||Ax - b||^2 + \lambda ||x||_1, \tag{45}$$

where $A \in \mathbb{R}^{n \times d}$ with n < d. The ground truth x^* is sparse, supported on 50 randomly selected indices. Each nonzero entry is drawn as $x_i = \eta_i + \mathrm{sgn}(\eta_i)$, with $\eta_i \sim \mathcal{N}(0,1)$. The design matrix A is Gaussian with column-wise variance scaling: $a_{ij} \sim \mathcal{N}(0,j^2)$. The response vector is generated as $b = Ax^* + \xi$, where $\xi \sim \mathcal{N}(0,\sigma^2I_n)$ with $\sigma = 1$. We set $\lambda = 1$ throughout the experiments.

We adopt the reference function $\phi(x) = \frac{1}{2}x^{T}Dx$, where $D = \operatorname{diag}(A^{T}A)$. Since $\nabla^{2}f = A^{T}A$, the smoothness constant is $L_{f} = \rho(A^{T}A)$, and the relative smoothness constant is $L = \rho(D^{-1/2}A^{T}AD^{-1/2})$. As the problem is over-parameterized (n < d), the relative strong convexity vanishes $(\mu = 0)$, and we employ the perturbed variant of Acc-MD; see Algorithm 2 in Appendix A.

As D is diagonal and positive definite, the Acc-MD subproblem (25) can be solved by a generalized soft-thresholding operator in closed form.

As shown in Fig. 2 and Fig. 3, the proposed Acc-MD method converges significantly faster than the competing approaches.



-Acadd -Acadd -Acadd -Abrid -A

Figure 2: Function error curve of LASSO problem, n = 150, d = 200.

Figure 3: Function error curve of LASSO problem, n = 1000, d = 2000.

Quadratic on the simplex We consider the constrained optimization problem over the (n-1)-dimensional simplex $\Delta_n = \{x \in \mathbb{R}^n \mid \sum_{i=1}^n x_i = 1, \ x_i \geq 0\}$:

$$\min_{x \in \Delta_n} \ f(x) := \frac{1}{2} \|Ax - b\|^2.$$

This can be viewed as a composite optimization problem, where g is the indicator function of Δ_n .

Let $A=(a_{ij})\in\mathbb{R}^{n\times n}$ be a Gaussian matrix with $a_{ij}\sim\mathcal{N}(0,j^2)$, and set the ground truth to $x^*=(1,\ldots,1)^\top/n$ with $b=Ax^*$. We use the reference function $\phi(x)=\frac{1}{2}x^\top Dx$, where $D=\operatorname{diag}(A^\top A)$ is strictly positive definite. The mirror maps $\nabla\phi$ and $\nabla\phi^*$ are then simple to evaluate, and the relative smoothness and convexity constants L,μ can be estimated analogously. As shown in Table 1, L remains nearly constant across problem sizes, and the relative condition number L/μ is significantly better than the Euclidean counterpart L_f/μ_f .

The Acc-MD subproblem (25) becomes

$$x_{k+1} \in \arg\min_{x \in \Delta_n} \|x - z_{k+1}\|_D^2, \quad \text{with } z_{k+1} = D^{-1}\tilde{z}_{k+1},$$
 (46)

where $\tilde{z}_{k+1} = \frac{1}{1+\alpha} \left[\nabla \phi(y_k) + \alpha \nabla \phi(x_k) - \frac{\alpha}{\mu} \nabla f(x_k) \right]$. Problem (46) is a projection onto the simplex under a weighted norm. We solve it by introducing the Lagrangian:

$$\mathcal{L}(x,\lambda,\beta) = \frac{1}{2} \|x - z\|_D^2 - \lambda \left(\sum_{i=1}^n x_i - 1\right) - \beta^\top x,$$

with KKT conditions:

$$d_i(x_i - z_i) - \lambda - \beta_i = 0, \quad x_i \ge 0, \quad \beta_i \ge 0, \quad x_i \beta_i = 0, \quad 1 \le i \le n, \quad \sum_{i=1}^n x_i = 1.$$

The solution satisfies $x_i = \max(0, z_i + \lambda/d_i)$, and λ is the unique solution to

$$\sum_{i=1}^{n} \max\left(0, z_i + \frac{\lambda}{d_i}\right) = 1,$$

which can be efficiently found by the bisection method.

As shown in Table 1, Acc-MD outperforms competing methods in both iteration count and runtime. The non-accelerated BPG method is omitted as it fails to converge within a reasonable time.

Table 1: Quadratic optimization on the simplex. Stopping criterion: $f(x) < 10^{-12} f(x_0)$.

Problem Size					Acc-MD		ABPG		FISTA		AOR-HB	
\overline{n}	L_f	μ_f	L	μ	#Iter	Time	#Iter	Time	#Iter	Time	#Iter	Time
125	4.04×10^6	0.6228	3.78	9.86×10^{-5}	1159	0.16	5225	0.47	39145	1.14	22930	0.53
250	3.17×10^{7}	0.0146	3.96	1.99×10^{-7}	1669	0.38	6602	1.20	81633	5.25	24133	1.68
500				3.79×10^{-7}				3.18				14.49
1000	2.00×10^{9}	0.0592	3.95	1.34×10^{-7}	4960	6.07	6221	10.26	118364	177.83	100283	129.71

LLM USAGE

In preparing this manuscript, large language models (LLMs) were employed exclusively to assist with language-related tasks, such as improving readability, grammar, and style. The models were not used for research ideation, development of methods, data analysis, or interpretation of results. All scientific content, including problem formulation, theoretical analysis, and experimental validation, was conceived, executed, and verified entirely by the authors. The authors bear full responsibility for the accuracy and integrity of the manuscript.

ETHICS STATEMENT

This work is purely theoretical and algorithmic, focusing on convex optimization methods. It does not involve human subjects, sensitive data, or applications that raise ethical concerns related to privacy, security, fairness, or potential harm. All experiments are based on publicly available datasets or synthetic data generated by standard procedures. The authors believe that this work fully adheres to the ICLR Code of Ethics.

REPRODUCIBILITY STATEMENT

We have taken several measures to ensure the reproducibility of our results. All theoretical assumptions are explicitly stated, and complete proofs are provided in the appendix. For the experimental evaluation, we describe the setup, parameter choices, and baselines in detail in the main text. The source code for our algorithms and experiments are available as supplementary materials. Together, these resources should allow others to reproduce and verify our theoretical and empirical findings.