Bregman proximal augmented Lagrangian revisited: Relative errors, ergodic convergence and efficient first-order oracle

Emanuel Laude* Panagiotis Patrinos*

April 19, 2023

Abstract

In this paper we study the convergence of the Bregman proximal augmented Lagrangian method with relative errors. We refine existing analyses for the Bregman proximal point and Bregman augmented Lagrangian method by means of asymptotic dual convergence and primal convergence in the ergodic sense. As a byproduct we show ergodic convergence rates of arbitrary order depending on the choice of the step-size. In addition, we discuss the efficient solution of the subproblems for linear inequality constraints using a recent nonlinear dual space preconditioning scheme for the proximal gradient method called anisotropic proximal gradient. The algorithm is compared to the primal-dual hybrid-gradient method on the task of linear programming with inequality constraints.

Keywords: augmented Lagrangian · duality · Bregman distance · nonlinear preconditioning · proximal point

AMS Subject Classification: 65K05 · 49J52 · 90C30

1 Introduction

In this paper we revisit the Bregman augmented Lagrangian method with proximal regularization. The method is a special case of the Bregman proximal point algorithm for solving monotone inclusions with relative errors due to Solodov and Svaiter [6]. Since Bregman distances, in general, do not have full domain, existing analyses [6] of the Bregman proximal point algorithm are too general to fully explain primal and dual convergence as well as asymptotic feasibility. In this paper we refine the existing analysis of [6] by means of asymptotic dual convergence as well as asymptotic feasiblity and primal convergence in the ergodic sense. [3] considers the Bregman augmented Lagrangian method with absolute errors. Following [3], we adopt Rockafellars abstract perturbation framework for convex duality [5, Chapter 11H]. This allows us to cover more general convex problems beyond equality and inequality constraints as in the classical exponential method of multipliers [7].

If not stated otherwise we adopt the notation from [5] and the notion of a Legendre function from [1].

*KU Leuven, Department of Electrical Engineering (ESAT-STADIUS), Kasteelpark Arenberg 10, 3001 Leuven, Belgium {emanuel.laude,panos.patrinos}@esat.kuleuven.be

2 Preliminaries

In this section we recapitulate the perturbation framework of Rockafellar for convex duality, see [5, Chapter 11H]. Let $F : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ be proper, lsc and convex. We define the primal problem as

$$P(0) := \inf_{x \in \mathbb{R}^n} F(x, 0). \tag{1}$$

Under appropriate conditions, see [5, Theorem 11.39], we have that $P(0) = P^{**}(0)$, where the latter gives rise to the dual problem.

We compute

$$\begin{aligned} P^*(y) &= \sup_{u \in \mathbb{R}^m} \langle u, y \rangle - P(u) \\ &= \sup_{(x,u) \in \mathbb{R}^n \times \mathbb{R}^m} \langle x, 0 \rangle + \langle u, y \rangle - F(x, u) \\ &= F^*(0, y) \end{aligned}$$

Then

$$P^{**}(0) = \sup_{y \in \mathbb{R}^m} -P^*(y) = \sup_{y \in \mathbb{R}^m} -F^*(0, y) = -\inf_{y \in \mathbb{R}^m} F^*(0, y) =: Q(0),$$

where $\sup_{y \in \mathbb{R}^m} -F^*(0, y) = Q(0)$ is the dual problem. Define the Lagrangian at (x, y) as the negative partial conjugate of $F(x, \cdot)$ at y:

$$L(x, y) := \inf_{u \in \mathbb{R}^m} F(x, u) - \langle y, u \rangle. \tag{2}$$

In this context F is also called a dualizing parametrization of the primal problem $\inf P$; see [5, Definition 11.45]. Consider the maximally monotone saddle-point operator $T : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^n \times \mathbb{R}^m$.

$$T(x, y) = \partial_x L(x, y) \times \partial_y (-L)(x, y). \tag{3}$$

Then thanks to [5, Proposition 11.48] we have that

$$(v, y) \in \partial F(x, u) \Leftrightarrow (v, u) \in T(x, y) \Leftrightarrow (x, u) \in \partial F^*(v, y). \tag{4}$$

A running example considered in this paper are linear inequality constraints:

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad Ax \leq b, \tag{5}$$

for $f \in \Gamma_0(\mathbb{R}^n)$ and $A \in \mathbb{R}^{m \times n}$. We consider the dualizing parametrization $F : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$:

$$F(x, u) := \begin{cases} f(x) & \text{if } Ax + u \leq b \\ +\infty & \text{otherwise.} \end{cases} \tag{6}$$

Then the Lagrangian takes the form

$$L(x, y) := \inf_{u \in \mathbb{R}^m} F(x, u) - \langle y, u \rangle \tag{7}$$

$$= \begin{cases} f(x) + \langle y, Ax - b \rangle & \text{if } y \geq 0 \\ -\infty & \text{otherwise,} \end{cases} \tag{8}$$

which is the classical augmented Lagrangian. The saddle-point operator (3) then amounts to:

$$T(x,y) = \partial_x L(x,y) \times \partial_y(-L)(x,y) \tag{9}$$
$$= \partial f(x) + A^\top y \times -(Ax-b) + N_{\mathbb{R}^m_+}(y). \tag{10}$$

For the remainder of this paper we assume that strong duality $P(0) = Q(0)$ holds and the conditions in [5, Theorem 11.39] are met to guarantee existence of primal and dual solutions.

3 Bregman proximal augmented Lagrangian with relative errors

To begin with we consider the Bregman proximal point algorithm for solving (1). We define the Bregman distance generated by Legendre function $\Phi \in \Gamma_0(\mathbb{R}^n \times \mathbb{R}^m)$:

$$D_\Phi(x,y) := \begin{cases} \Phi(x) - \Phi(y) - \langle \nabla\Phi(y), x-y\rangle & x \in \operatorname{dom}\Phi, y \in \operatorname{int}\operatorname{dom}\Phi \\ +\infty & \text{otherwise.} \end{cases} \tag{11}$$

In addition we require Φ to comply with the following properties adopted from [6, Definition 2.1]. Some of which are redundant with regards to the Legendre property of Φ; see [1].

Definition 3.1 (Bregman function)**.** *Let $\Phi \in \Gamma_0(\mathbb{R}^n \times \mathbb{R}^m)$. Then we say that Φ is Bregman if the following properties hold true:*

(i) *Φ is continuous and striclty convex on* $\operatorname{dom}\Phi$.

(ii) *Φ is continuously differentiable on* $\operatorname{int}\operatorname{dom}\Phi$.

(iii) *for any $x \in \operatorname{dom}\Phi$ and $\alpha \in \mathbb{R}$ the right partial level-set $\{y \in \operatorname{dom}\Phi : D_\Phi(x,y) \le \alpha\}$ is bounded.*

(iv) *For $\{y^k\}_{k=0}^\infty \subset \operatorname{int}\operatorname{dom}\Phi$ such that $y^k \to y$ it holds that $\lim_{k\to\infty} D_\Phi(y, y^k) \to 0$.*

We consider the monotone mapping $T : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^n \times \mathbb{R}^m$ with the property that $\operatorname{int}\operatorname{dom}\Phi \cap \operatorname{dom} T \neq \emptyset$ and the sequence of step-sizes $\{\sigma_k\}_{k=0}^\infty$ that are bounded away from zero for all k. Then the algorithm for reference function Φ takes the form:

$$z^{k+1} = (\nabla\Phi + \sigma_k T)^{-1}(\nabla\Phi^*(z^k)). \tag{12}$$

To account for errors we consider an inexact variant of Bregman proximal point with relative errors due to [6] which is given as

$$w^{k+1} \in T(p^{k+1}) \tag{13}$$
$$z^{k+1} = \nabla\Phi^*(\nabla\Phi(z^k) - \sigma_k w^{k+1}) \tag{14}$$
$$D_\Phi(p^{k+1}, z^{k+1}) \le \rho D_\Phi(p^{k+1}, z^k). \tag{15}$$

We state the four-point lemma [6, Corollary 2.6]:

Lemma 3.2 (four-point lemma)**.** *Let $\Phi \in \Gamma_0(\mathbb{R}^n \times \mathbb{R}^m)$ be Bregman and Legendre. Let $x, z \in \operatorname{int}(\operatorname{dom}\Phi)$ and $y, s \in \operatorname{dom}\Phi$. Then we have that*

$$D_\Phi(s,z) = D_\Phi(s,x) + \langle \nabla\Phi(x) - \nabla\Phi(z), s-y\rangle + D_\Phi(y,z) - D_\Phi(y,x).$$

The following result is a direct consequence of [6, Lemma 4.1]. For completeness a simple proof is provided.

Lemma 3.3 (Fejér monotonicity)**.** *Let $z^\star \in T^{-1}(0)$ and assume that Φ is Bregman and Legendre. Let $0 \leq \rho < 1$. Then we have that*

$$\begin{aligned} D_\Phi(z^\star, z^{k+1}) &\leq D_\Phi(z^\star, z^k) - \langle \sigma_k w^{k+1} - 0, p^{k+1} - z^\star \rangle - (1-\rho) D_\Phi(p^{k+1}, z^k) \\ &\leq D_\Phi(z^\star, z^k) - (1-\rho) D_\Phi(p^{k+1}, z^k). \end{aligned}$$

Proof. We have thanks to Lemma 3.2 for $s = z^\star$, $z = z^{k+1}$, $x = z^k$ and $y = p^{k+1}$:

$$\begin{aligned} D_\Phi(z^\star, z^{k+1}) &= D_\Phi(z^\star, z^k) + \langle \nabla\Phi(z^k) - \nabla\Phi(z^{k+1}), z^\star - p^{k+1} \rangle + D_\Phi(p^{k+1}, z^{k+1}) - D_\Phi(p^{k+1}, z^k) \\ &= D_\Phi(z^\star, z^k) - \langle \sigma_k w^{k+1} - 0, p^{k+1} - z^\star \rangle + D_\Phi(p^{k+1}, z^{k+1}) - D_\Phi(p^{k+1}, z^k) \\ &\leq D_\Phi(z^\star, z^k) - \langle \sigma_k w^{k+1} - 0, p^{k+1} - z^\star \rangle - (1-\rho) D_\Phi(p^{k+1}, z^k) \\ &\leq D_\Phi(z^\star, z^k) - (1-\rho) D_\Phi(p^{k+1}, z^k) \end{aligned}$$

where the first equality follows from the extra-gradient step (14) in the definition of the algorithm and the last inequalities follow from the error bound (15) and the monotonicity of T respectively. □

Lemma 3.4. *Let $z^\star \in T^{-1}(0)$ and assume that $\Phi \in \Gamma_0(\mathbb{R}^n \times \mathbb{R}^m)$ is Bregman and Legendre. Let $0 \leq \rho < 1$. Then the following properties hold true:*

(i) $\{D_\Phi(z^\star, z^k)\}_{k=0}^\infty$ *is convergent.*

(ii) $\{D_\Phi(p^{k+1}, z^k)\}_{k=0}^\infty$ *and* $\{D_\Phi(p^{k+1}, z^{k+1})\}_{k=0}^\infty$ *are summable and in particular converge to* 0*.*

(iii) $\langle \sigma_k w^{k+1}, p^{k+1} - z^\star \rangle \to 0$*.*

(iv) $\{z^k\}_{k=0}^\infty$ *and* $\{p^k\}_{k=0}^\infty$ *are bounded.*

Proof. Summing the inequality from Lemma 3.3 we obtain that

$$0 \leq (1-\rho) \sum_{k=0}^{K} D_\Phi(p^{k+1}, z^k) \leq D_\Phi(z^\star, z^0) - D_\Phi(z^\star, z^K) \leq D_\Phi(z^\star, z^0), \tag{16}$$

and thus $\{D_\Phi(p^{k+1}, z^k)\}_{k=0}^\infty$ is summable implying that $D_\Phi(p^{k+1}, z^k) \to 0$. Since $D_\Phi(p^{k+1}, z^{k+1}) \leq \rho D_\Phi(p^{k+1}, z^k)$ we also have that $\{D_\Phi(p^{k+1}, z^{k+1})\}_{k=0}^\infty$ is summable and $D_\Phi(p^{k+1}, z^{k+1}) \to 0$. Thanks to the inequality from Lemma 3.3 we obtain further that $\{D_\Phi(z^\star, z^k)\}_{k=0}^\infty$ is monotoncially decreasing. By nonnegativity of the Bregman distance it is also convergent. This implies that $D_\Phi(z^\star, z^k) - D_\Phi(z^\star, z^{k+1}) \to 0$ and therefore

$$\langle \sigma_k w^{k+1}, p^{k+1} - z^\star \rangle = D_\Phi(z^\star, z^k) - D_\Phi(z^\star, z^{k+1}) + D_\Phi(p^{k+1}, z^{k+1}) - D_\Phi(p^{k+1}, z^k) \to 0,$$

as claimed. Since $D_\Phi(z^\star, \cdot)$ is level-bounded $\{z^k\}_{k=0}^\infty$ is bounded. Suppose that $\{p^k\}_{k=0}^\infty$ is unbounded. Take a subsequence indexed by k_j along which $\|p^{k_j}\| \to \infty$ for $k_j \to \infty$. Since z^{k_j} is bounded we can make sure that $z^{k_j} \to \bar{z}$ for some $\bar{z} \in \operatorname{dom}\Phi$ and $j \to \infty$. Since $D_\Phi(p^{k_j}, z^{k_j}) \to 0$, in light of [6, Theorem 2.4] we have $p^{k_j} \to \bar{z}$ a contradiction and thus $\{p^k\}_{k=0}^\infty$ are bounded. □

In spite of its generality, in the context of Bregman augmented Lagrangian, [6, Theorem 3.2] is only applicable if $z^\star \in \operatorname{int}(\operatorname{dom}\Phi)$. Thus we shall refine the analysis of [6] for the Bregman proximal augmented Lagrangian method.

More specifically, we choose $\psi \in \Gamma_0(\mathbb{R}^n)$ Legendre, super-coercive with full domain and $\phi \in \Gamma_0(\mathbb{R}^m)$ Legendre such that $D_\phi(x, \cdot)$ is level-bounded for all $x \in \operatorname{dom}\phi$. We define the separable Legendre function $\Phi(z) := \psi(x) + \phi(y)$ for $z = (x, y)$. Choose T the saddle-point operator corresponding to the subdifferential of the Lagrangian L as in (3). Then the Bregman proximal point algorithm (13)–(15) takes the form:

$$\begin{aligned}(v^{k+1}, u^{k+1}) &\in T(s^{k+1}, r^{k+1})\\ x^{k+1} &= \nabla\psi^*(\nabla\psi(x^k) - \sigma_k v^{k+1})\\ y^{k+1} &= \nabla\phi^*(\nabla\phi(y^k) - \sigma_k u^{k+1})\\ D_\psi(s^{k+1}, x^{k+1}) + D_\phi(r^{k+1}, y^{k+1}) &\le \rho D_\psi(s^{k+1}, x^k) + \rho D_\phi(r^{k+1}, y^k).\end{aligned}$$

In addition, the dual update is solved exactly whereas we allow for an error in the primal update: That is $(v^{k+1}, u^{k+1}) \in T(s^{k+1}, r^{k+1})$ with $r^{k+1} = y^{k+1}$ and thus $D_\phi(r^{k+1}, y^{k+1}) = 0$. Then the algorithm further specializes to

$$(v^{k+1}, u^{k+1}) \in T(s^{k+1}, y^{k+1}) \tag{17}$$
$$x^{k+1} = \nabla\psi^*(\nabla\psi(x^k) - \sigma_k v^{k+1}) \tag{18}$$
$$y^{k+1} = \nabla\phi^*(\nabla\phi(y^k) - \sigma_k u^{k+1}) \tag{19}$$
$$D_\psi(s^{k+1}, x^{k+1}) \le \rho D_\psi(s^{k+1}, x^k) + \rho D_\phi(y^{k+1}, y^k). \tag{20}$$

Invoking the Fejèr monotonicity property proved in Lemma 3.3 we obtain following result for the specific algorithm:

Lemma 3.5. *Let $\phi \in \Gamma_0(\mathbb{R}^n)$ and $\psi \in \Gamma_0(\mathbb{R}^m)$ be Bregman and Legendre. Assume that ψ is super-coercive with $\operatorname{dom}\psi = \mathbb{R}^n$. Let $z^\star = (x^\star, y^\star) \in T^{-1}(0)$. Then the following properties hold true:*

(i) *$D_\psi(x^\star, x^k) + D_\phi(y^\star, y^k)$ is convergent.*

(ii) *$D_\phi(y^{k+1}, y^k)$, $D_\psi(s^{k+1}, x^k)$ and $D_\psi(s^{k+1}, x^{k+1})$ are summable*

(iii) *$\{s^k\}_{k=0}^\infty$, $\{x^k\}_{k=0}^\infty$ and $\{y^k\}_{k=0}^\infty$ are bounded*

(iv) *$v^k \to 0$ and $\langle u^k, y^k - y^\star\rangle \to 0$.*

Proof. Recall that $p^k = (s^k, r^k)$, $z^k = (x^k, y^k)$ and $\Phi(z) := \psi(x) + \phi(y)$. Since $r^k = y^k$ we have $D_\Phi(p^{k+1}, z^k) = D_\psi(s^{k+1}, x^k) + D_\phi(y^{k+1}, y^k)$. Thus Lemma 3.4 yields the first three assertions. Since ψ is super-coercive with full domain $\operatorname{dom}\psi^* = \mathbb{R}^n$ and ψ^* is super-coercive too. Via continuity of $D_\psi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ and $\nabla\psi : \mathbb{R}^n \to \mathbb{R}^n$ and boundedness of $\{x^k\}_{k=0}^\infty$ and $\{s^k\}_{k=0}^\infty$ via the fact that $D_\psi(x, s) = 0$ if and only if $x = s$ and injectivity of $\nabla\psi$ we have that $\|\nabla\psi(s^{k+1}) - \nabla\psi(x^{k+1})\| \to 0$ as well as $\|\nabla\psi(s^{k+1}) - \nabla\psi(x^k)\| \to 0$. Thus we have that $0 \le \sigma_k\|v^{k+1}\| = \|\nabla\psi(x^{k+1}) - \nabla\psi(x^k)\| \le \|\nabla\psi(x^{k+1}) - \nabla\psi(s^{k+1})\| + \|\nabla\psi(s^{k+1}) - \nabla\psi(x^k)\| \to 0$ and since σ_k is bounded away from 0 we also have that $v^k \to 0$. Since x^k is bounded this implies that $\sigma_k\langle v^k, x^k - x^\star\rangle \to 0$. Recall that $w^k = (v^k, u^k)$, $z^k = (x^k, y^k)$ and $z^\star = (x^\star, y^\star)$. Thanks to Lemma 3.4 we have that

$$\sigma_k\langle w^k, z^k - z^\star\rangle = \sigma_k\langle v^k, x^k - x^\star\rangle + \sigma_k\langle u^k, y^k - y^\star\rangle \to 0.$$

and thus

$$\sigma_k \langle u^k, y^k - y^\star \rangle \to 0$$

and since σ_k is bounded away from 0 we have that $\langle u^k, y^k - y^\star \rangle \to 0$. □

Unfortunately, the sequence u^k is hard to control, unless the limit point lies in $\operatorname{int} \operatorname{dom} \phi$. Therefore Fejér montonicity cannot be used directly to establish convergence. Instead, the following result establishes the dual asymptotic convergence using the dual perturbation function F^*:

Theorem 3.6 (dual asymptotic convergence)**.** *For any dual solution $y^\star \in \arg\max -F^*(0, \cdot)$ we have that $F^*(v^k, y^k) \to F^*(0, y^\star)$. In addition every limit point y^∞ of the sequence of dual iterates $\{y^k\}_{k=0}^\infty$ is a dual solution, i.e., $y^\infty \in \arg\max -F^*(0, \cdot)$.*

Proof. In light of (4) $(v^k, u^k) \in T(s^k, y^k)$ if and only if $(s^k, u^k) \in \partial F^*(v^k, y^k)$. We have that

$$F^*(0, y^\star) \geq F^*(v^k, y^k) + \langle s^k, 0 - v^k \rangle + \langle u^k, y^\star - y^k \rangle,$$

which we rearrange to

$$F^*(v^k, y^k) \leq F^*(0, y^\star) + \langle s^k, v^k \rangle - \langle u^k, y^\star - y^k \rangle. \tag{21}$$

In light of Lemma 3.5 we have that $v^k \to 0$, $\{s^k\}_{k=0}^\infty$ is bounded and $\langle u^k, y^\star - y^k \rangle \to 0$. Then (21) yields that

$$\limsup_{k \to \infty} \; F^*(v^k, y^k) \leq F^*(0, y^\star).$$

We now consider $\liminf_{k \to \infty} \; F^*(v^k, y^k)$ which by the above is at most $F^*(0, y^\star)$. Consider a subsequence indexed by k_j which realizes the lim inf which is possibly $-\infty$:

$$\lim_{j \to \infty} \; F^*(v^{k_j}, y^{k_j}) = \liminf_{k \to \infty} \; F^*(v^k, y^k).$$

Since y^{k_j} is bounded we can consider a subsequence if necessary to make sure that $y^{k_j} \to y^\infty$. Since $y^\star \in \arg\min F^*(0, \cdot)$ we have

$$\begin{aligned} F^*(0, y^\star) &\leq F^*(0, y^\infty) \\ &\leq \liminf_{j \to \infty} F^*(v^{k_j}, y^{k_j}) \\ &= \liminf_{k \to \infty} \; F^*(v^k, y^k), \end{aligned}$$

where the second inequality follows by lower semi-continuity and the last equality by the choice of the subsequence with indexes k_j. Hence we have

$$\lim_{k \to \infty} \; F^*(v^k, y^k) = F^*(0, y^\star).$$

Let y^∞ be any limit point of the sequnce $\{y^k\}_{k=0}^\infty$. Consider $y^{k_j} \to y^\infty$. Then we have $F^*(0, y^\star) \leq F^*(0, y^\infty) \leq \liminf_{j \to \infty} F^*(v^{k_j}, y^{k_j}) = F^*(0, y^\star) = \inf_{y \in \mathbb{R}^m} F^*(0, y)$ and thus y^∞ is a dual solution, i.e., $y^\infty \in \arg\max -F^*(0, \cdot)$. □

Unfortunately, primal convergence cannot be obtained in the Bregman case unless y^∞ lies in $\operatorname{int} \operatorname{dom} \phi$ in which case we can invoke [6, Theorem 3.2(1)].

Instead we show ergodic convergence in the primal.

Recall the celebrated three-point lemma [2, Lemma 3.1]:

Lemma 3.7 (three-point lemma)**.** *Let $\phi \in \Gamma_0(\mathbb{R}^n)$ be Bregman and Legendre. Then for any $x, y \in \operatorname{int}(\operatorname{dom}\phi)$ and $z \in \operatorname{dom}\phi$ the following identity holds true:*

$$D_\phi(z,y) = D_\phi(x,y) + D_\phi(z,x) + \langle \nabla\phi(x) - \nabla\phi(y), z - x\rangle. \tag{22}$$

Lemma 3.8. *We have for any $x \in \mathbb{R}^n$ and any $y \in \operatorname{dom}\phi$*

$$L(s^{k+1}, y) - L(x, y^{k+1}) \leq \tfrac{1}{\sigma_k}\big(D_\psi(x,x^k) + D_\phi(y,y^k) - D_\psi(x,x^{k+1}) - D_\phi(y,y^{k+1})\big),$$

Proof. We have that $v^{k+1} \in \partial_x L(s^{k+1}, y^{k+1})$. By convexity of L in x we have for all $x \in \mathbb{R}^n$ that

$$L(x, y^{k+1}) \geq L(s^{k+1}, y^{k+1}) + \langle v^{k+1}, x - s^{k+1}\rangle. \tag{23}$$

Using (18) this means equivalently that

$$\begin{aligned} L(s^{k+1}, y^{k+1}) - L(x, y^{k+1}) &\leq \langle v^{k+1}, s^{k+1} - x\rangle \\ &= \tfrac{1}{\sigma_k}\langle \nabla\psi(x^k) - \nabla\psi(x^{k+1}), s^{k+1} - x\rangle. \end{aligned} \tag{24}$$

We have thanks to Lemma 3.2 for $s = x$, $z = x^{k+1}$, $x = x^k$ and $y = s^{k+1}$:

$$\begin{aligned} \langle \nabla\psi(x^k) - \nabla\psi(x^{k+1}), s^{k+1} - x\rangle &= D_\psi(x,x^k) - D_\psi(x,x^{k+1}) + D_\psi(s^{k+1},x^{k+1}) - D_\psi(s^{k+1},x^k) \\ &\leq D_\psi(x,x^k) - D_\psi(x,x^{k+1}) - (1-\rho)D_\psi(s^{k+1},x^k) \\ &\quad + \rho D_\phi(y^{k+1},y^k), \end{aligned} \tag{25}$$

where the last inequality follows from the error bound (20). Combining the estimates (24) and (25) we have

$$\begin{aligned} L(s^{k+1}, y^{k+1}) - L(x, y^{k+1}) &\leq \langle v^{k+1}, s^{k+1} - x\rangle \\ &\leq \tfrac{1}{\sigma_k}\big(D_\psi(x,x^k) - D_\psi(x,x^{k+1}) - (1-\rho)D_\psi(s^{k+1},x^k) \\ &\quad + \rho D_\phi(y^{k+1},y^k)\big). \end{aligned} \tag{26}$$

We have that $u^{k+1} \in \partial_y(-L)(s^{k+1}, y^{k+1})$. By convexity of $-L$ in y we have for any $y \in \mathbb{R}^m$

$$-L(s^{k+1}, y) \geq -L(s^{k+1}, y^{k+1}) + \langle u^{k+1}, y - y^{k+1}\rangle.$$

Using (19) this means equivalently that

$$-L(s^{k+1}, y) \geq -L(s^{k+1}, y^{k+1}) + \tfrac{1}{\sigma_k}\langle \nabla\phi(y^k) - \nabla\phi(y^{k+1}), y - y^{k+1}\rangle. \tag{27}$$

By Lemma 3.7 we have that

$$\langle y - y^{k+1}, \nabla\phi(y^{k+1}) - \nabla\phi(y^k)\rangle = D_\phi(y,y^k) - D_\phi(y^{k+1},y^k) - D_\phi(y,y^{k+1}) \tag{28}$$

and thus

$$L(s^{k+1}, y) - L(s^{k+1}, y^{k+1}) \leq \tfrac{1}{\sigma_k}\big(D_\phi(y,y^k) - D_\phi(y^{k+1},y^k) - D_\phi(y,y^{k+1})\big). \tag{29}$$

Summing (26) and (29) we obtain

$$\begin{aligned} L(s^{k+1}, y) - L(x, y^{k+1}) &\leq \tfrac{1}{\sigma_k}\big(D_\psi(x,x^k) - D_\psi(x,x^{k+1}) + D_\phi(y,y^k) - D_\phi(y,y^{k+1}) \\ &\quad - (1-\rho)(D_\psi(s^{k+1},x^k) + D_\phi(y^{k+1},y^k))\big) \\ &\leq \tfrac{1}{\sigma_k}\big(D_\psi(x,x^k) - D_\psi(x,x^{k+1}) + D_\phi(y,y^k) - D_\phi(y,y^{k+1})\big) \end{aligned}$$

as claimed. □

We define the ergodic iterates:

$$\breve{s}^K := \frac{\sum_{k=0}^{K-1} \sigma_k s^{k+1}}{\sum_{k=0}^{K-1} \sigma_k} \quad \text{and} \quad \breve{y}^K := \frac{\sum_{k=0}^{K-1} \sigma_k y^{k+1}}{\sum_{k=0}^{K-1} \sigma_k}. \tag{30}$$

Theorem 3.9. *We have for any $x \in \mathbb{R}^n$ and any $y \in \operatorname{dom} \phi$:*

$$L(\breve{s}^K, y) - L(x, \breve{y}^K) \leq \frac{1}{\sum_{k=0}^{K-1} \sigma_k} \big(D_\psi(x, x^0) + D_\phi(y, y^0) \big).$$

In particular any limit point $\breve{s}^\infty$ of the sequence of primal ergodic iterates $\breve{s}^K$ is a primal solution, i.e.,

$$\inf F(\cdot, 0) = F(\breve{s}^\infty, 0)$$

and any limit point of the sequence of dual iterates $\breve{y}^\infty$ is a dual solution, i.e.,

$$\sup -F^*(0, \cdot) = -F^*(0, \breve{y}^\infty).$$

Proof. Since L is convex in x and concave in y we have in view of Lemma 3.8 that

$$\begin{aligned}
L(\breve{s}^K, y) - L(x, \breve{y}^K) &\leq \frac{1}{\sum_{k=0}^{K-1} \sigma_k} \sum_{k=0}^{K-1} \sigma_k \big(L(s^{k+1}, y) - L(x, y^{k+1}) \big) \\
&\leq \frac{1}{\sum_{k=0}^{K-1} \sigma_k} \sum_{k=0}^{K-1} \big(D_\psi(x, x^k) + D_\phi(y, y^k) - D_\psi(x, x^{k+1}) - D_\phi(y, y^{k+1}) \big) \\
&= \frac{1}{\sum_{k=0}^{K-1} \sigma_k} \big(D_\psi(x, x^0) + D_\phi(y, y^0) - D_\psi(x, x^K) - D_\phi(y, y^K) \big) \\
&\leq \frac{1}{\sum_{k=0}^{K-1} \sigma_k} \big(D_\psi(x, x^0) + D_\phi(y, y^0) \big).
\end{aligned}$$

No consider a convergent subsequence $\breve{s}^{K_j} \to \breve{s}^\infty$. By going to another subsequence if necessary we can ensure that $\breve{y}^{K_j} \to \breve{y}^\infty$. Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. Since $L(\cdot, y)$ and $-L(x, \cdot)$ are both lsc we have that

$$L(\breve{s}^\infty, y) - L(x, \breve{y}^\infty) \leq \liminf_{j \to \infty} L(\breve{s}^{K_j}, y) - L(x, \breve{y}^{K_j}) \leq 0.$$

This implies that $(\breve{s}^\infty, \breve{y}^\infty)$ is a saddle-point of L. In light of [5, Theorem 11.50] $\breve{s}^\infty$ is a primal solution. The argument for the dual solution follows similarly. □

In what follows we specialize to problems with linear inequality constraints as in (5).

The following theorem is a refinement of Theorem 3.9 showing asymptotic feasibility for linearly constrained problems. To this end we adapt the proof of [8, Theorem 3.1].

Theorem 3.10. *We have that*

$$\max\{|f(\breve{s}^K) - f(x^\star)|, \|[A\breve{s}^K - b]_+\|\} \leq \big(D_\psi(x^\star, x^0) + \max_{y \in \mathbb{R}^m_+ : \|y\| \leq 2\|y^\star\| + 1} D_\phi(y, y^0) \big)$$

Proof. Let $y \in \mathbb{R}^m_+$. Thanks to Theorem 3.9 we have for $x = x^\star$ that $Ax^\star - b = 0$ and thus

$$\begin{aligned} L(\breve{s}^K, y) - L(x^\star, \breve{y}^K) &= f(\breve{s}^K) - f(x^\star) + \langle y, A\breve{s}^K - b\rangle + \langle y^{k+1}, b - Ax^\star\rangle \\ &= f(\breve{s}^K) - f(x^\star) + \langle y, A\breve{s}^K - b\rangle. \end{aligned}$$

As a result we have

$$f(\breve{s}^K) - f(x^\star) + \langle y, A\breve{s}^K - b\rangle \leq \big(D_\psi(x^\star, x^0) + D_\phi(y, y^0)\big).$$

For $y = 0$ we obtain that

$$f(\breve{s}^K) - f(x^\star) \leq \big(D_\psi(x^\star, x^0) + D_\phi(0, y^0)\big).$$

Taking the maximum over $y \in \mathbb{R}^m_+$ with $\|y\| \leq 2\|y^\star\| + 1$ we obtain

$$f(\breve{s}^K) - f(x^\star) + (2\|y^\star\| + 1)\|[A\breve{s}^K - b]_+\| \leq \big(D_\psi(x^\star, x^0) + \max_{y \in \mathbb{R}^m_+ : \|y\| \leq 2\|y^\star\|+1} D_\phi(y, y^0)\big). \tag{31}$$

Since $(x^\star, y^\star)$ is a saddle-point of L we have for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m_+$ that

$$L(x, y^\star) - L(x^\star, y) \geq 0$$

and thus for $y = 0$ and $x = \breve{s}^K$

$$f(x^\star) - f(\breve{s}^K) - \|y^\star\|\|[A\breve{s}^K - b]_+\| \leq f(x^\star) - f(\breve{s}^K) - \langle y^\star, A\breve{s}^K - b\rangle \leq 0. \tag{32}$$

Summing (31) and (32) we obtain

$$\|[A\breve{s}^K - b]_+\| \leq (\|y^\star\| + 1)\|[A\breve{s}^K - b]_+\| \leq \big(D_\psi(x^\star, x^0) + \max_{y \in \mathbb{R}^m_+ : \|y\| \leq 2\|y^\star\|+1} D_\phi(y, y^0)\big). \tag{33}$$

The result then follows from the fact that $f(x^\star) - f(\breve{s}^K) \leq \|y^\star\|\|[A\breve{s}^K - b]_+\|$ □

3.1 Linearly convergent firt-order orcale

In this section we consider first-order oracles for the primal update for the linearly constrained problem (5). To this end we choose $\phi(x) = \sum_{j=1}^m x_j \log(x_j) - x_j$ being the entropy. For $\sigma_k = \sigma > 0$ constant and T as in (9) the exact Bregman proximal point iteration (12) takes the form

$$0 \in \partial f(x^{k+1}) + A^\top y^{k+1} + \sigma\nabla\psi(x^{k+1}) - \sigma\nabla\psi(x^k) \tag{34}$$

$$0 = -Ax^{k+1} + b + \sigma\nabla\phi(y^{k+1}) - \sigma\nabla\phi(y^k), \tag{35}$$

which corresponds to the first-order necessary and sufficient optimality condition of the saddle-point problem

$$(x^{k+1}, y^{k+1}) \in \operatorname*{arg\,minimax}_{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m} \; f(x) + \langle Ax - b, y\rangle - \sigma D_\phi(y, y^k) + \sigma D_\psi(x, x^k). \tag{36}$$

Equivalently this can be written in terms of

$$x^{k+1} \in \operatorname*{arg\,min}_{x\in\mathbb{R}^n} \; f(x) + \sigma \star \phi^*(Ax - b + \sigma\nabla\phi(y^k)) + \sigma D_\psi(x, x^k) \tag{37}$$

$$y^{k+1} = \nabla\phi^*(\nabla\phi(y^k) + \sigma^{-1}(Ax^{k+1} - b)). \tag{38}$$

In what follows we discuss the solution of the subproblem (37) with the anisotropic proximal gradient algorithm [4] with reference function ϕ^* which potentially attains linear convergence. For that purpose $\sigma D_\psi(\cdot, x^k)$ must be strongly convex in the anisotropic sense [4, Definition 5.8] relative to ϕ^*.

A simple choice for ψ is the reference function ϕ^* itself. However, ϕ^* is the exponential and as such $D_{\phi^*}(x, \cdot)$ is not level-bounded. Note that a convergent proximal point method requires this property to hold so as to ensure boundedness of the iterates.

Thus we adopt the following piecewise defined choice for ψ as to make sure the property holds true:

$$\psi(x) = \sum_{i=1}^{n} \ell(x_i) \quad \ell(t) = \begin{cases} \frac{1}{2}t^2 + t + 1 & \text{if } t < 0 \\ \exp(t) & \text{otherwise.} \end{cases} \tag{39}$$

Then ψ is strictly convex, continuously differentiable and super-coercive.

For attaining linear convergence with anisotropic proximal gradient we have to write the cost of the subproblem

$$\min_{x\in\mathbb{R}^n} \; f(x) + \sigma \star \phi^*(Ax - b + \sigma\nabla\phi(y^k)) + \sigma D_\psi(x, x^k)$$

in terms of the sum $h(x)+g(x)$ of an anisotropically smooth function h and an anisotropically strongly convex function g. Then we perform anisotropic forward steps with reference function ϕ^* on the anisotropically smooth part h and backward steps on the anisotropically strongly convex part g. Thanks to [4, Corollary 6.5] linear convergence will be attained. As shown in [4, Example 4.10],

$$h(x) = \sigma \star \phi^*(Ax - b + \sigma\nabla\phi(y^k))$$

is anisotropically smooth.

In most cases f will be linear. Then $f + \sigma D_\psi(\cdot, x^k)$ can be written as the sum of $\sigma\psi$ and a linear term $f - \sigma\langle\nabla\psi(x^k), \cdot\rangle$. Due to separability verfication of strong convexity boils down to strong convexity of

$$(\sigma\ell)(t) - at.$$

In order to show strong convexity of $\sigma\ell(t) - at$ we have to make sure that $a > 0$. This can be achieved as follows: If $a \leq 0$ the term $-at$ is anisotropically smooth relative to exp and thus we can add those terms $-at$ to the smooth part h preserving its anistropic smoothness property.

In order to show anisotropic strong convexity of $\sigma\ell(t) - at$ for $a > 0$ we invoke the dual characterization of anisotropic strong convexity and instead show that $(\sigma\star\ell^*)(s+a)$ is smooth in the Bregman sense relative to the entropy, i.e., $L(s\log(s) - s) - (\sigma \star \ell^*)(s + a)$ is convex on $(0, +\infty)$ for some $L > 0$.

To this end we compute the convex conjugate of ℓ:

$$\ell^*(s) = \sup_t ts - \ell(t). \tag{40}$$

We have

$$\ell'(t) = \begin{cases} t+1 & \text{if } t < 0 \\ \exp(t) & \text{otherwise.} \end{cases}$$

Equating the derivative to zero yields: $s = \ell'(t)$ and thus the conjugate translates to

$$t^*(s) = \begin{cases} s-1 & \text{if } s < 1 \\ \log(s) & \text{otherwise} \end{cases} \tag{41}$$

We insert this recurrence into

$$\ell^*(s) = t^*(s)s - \ell(t^*(s)) = \begin{cases} (s-1)s - \frac{1}{2}(s-1)^2 - (s-1) - 1 & \text{if } s < 1 \\ s\log(s) - s & \text{otherwise} \end{cases} \tag{42}$$

$$= \begin{cases} \frac{1}{2}s^2 - s - \frac{1}{2} & \text{if } s < 1 \\ s\log(s) - s & \text{otherwise} \end{cases} \tag{43}$$

We compute:

$$(\ell^*)''(s) = \begin{cases} 1 & \text{if } s < 1 \\ \frac{1}{s} & \text{otherwise} \end{cases}$$

Then we compute the second-order derivative of $L(s\log(s) - s) - (\sigma \star \ell^*)(s+a)$:

$$\begin{aligned} \tfrac{L}{s} - \big((\sigma \star \ell^*)(\cdot + a)\big)''(s) &= \tfrac{L}{s} - (\sigma \star \ell^*)''(s+a) \\ &= \tfrac{L}{s} - \tfrac{1}{\sigma}(\ell^*)''(\sigma^{-1}(s+a)) \\ &= \begin{cases} \frac{L}{s} - \frac{1}{\sigma} & \text{if } s + a < \sigma \\ \frac{L}{s} - \frac{1}{s+a} & \text{otherwise} \end{cases} \end{aligned}$$

This shows that for $L = 1$ we have that $L(s\log(s) - s) - \sigma \star \ell^*(s+a)$ is convex on $(0, +\infty)$. This means that $(\sigma \star \ell^*)(\cdot + a)$ is 1-Bregman smooth relative to $s \mapsto s\log(s) - s$. Dually, $\sigma\ell(t) - at$ is 1-anisotropically strongly convex relative to $\exp(t)$.

4 Conclusion

In this work we have considered the proximal Bregman augmented Lagrangian method with relative errors which is a special case of the Bregman proximal point algorithm with relative errors due to Solodov and Svaiter. We provided a dual asymptotic convergence analysis and primal asymptotic convergence and feasiblity in the ergodic sense. In addition, the efficient solution with of the subproblem for inequality constraints is discussed using a recent linealry convergent nonlinear preconditioning scheme for the proximal gradient method.

References

[1] H. H. Bauschke and J. M. Borwein. "Legendre functions and the method of random Bregman projections". In: *Journal of Convex Analysis* 4.1 (1997), pp. 27–67.

[2] G. Chen and M. Teboulle. "Convergence analysis of a proximal-like minimization algorithm using Bregman functions". In: *SIAM Journal on Optimization* 3.3 (1993), pp. 538–543.

[3] J. Eckstein. "A practical general approximation criterion for methods of multipliers based on Bregman distances". In: *Mathematical Programming* 96 (2003), pp. 61–86.

[4] E. Laude and P. Patrinos. "Anisotropic Proximal Gradient". In: *arXiv preprint arXiv:2210.15531* (2022).

[5] R. T. Rockafellar and R. J. Wets. *Variational Analysis*. New York: Springer, 1998.

[6] M. V. Solodov and B. F. Svaiter. "An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions". In: *Mathematics of Operations Research* 25.2 (2000), pp. 214–230.

[7] P. Tseng and D. P. Bertsekas. "On the convergence of the exponential multiplier method for convex programming". In: *Mathematical programming* 60.1-3 (1993), pp. 1–19.

[8] S. Yan and N. He. "Bregman Augmented Lagrangian and its acceleration". In: *arXiv preprint arXiv:2002.06315* (2020).