



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Tam, MK;Uteda, DJ

Title:

Bregman-Golden Ratio Algorithms for Variational Inequalities

Date:

2023-12

Citation:

Tam, M. K. & Uteda, D. J. (2023). Bregman-Golden Ratio Algorithms for Variational Inequalities. *Journal of Optimization Theory and Applications*, 199 (3), pp.993-1021.
<https://doi.org/10.1007/s10957-023-02320-2>.

Persistent Link:

<https://hdl.handle.net/11343/338155>

License:

[cc-by](#)



Bregman-Golden Ratio Algorithms for Variational Inequalities

Matthew K. Tam¹ · Daniel J. Uteda¹ 

Received: 29 July 2022 / Accepted: 26 September 2023
© The Author(s) 2023

Abstract

Variational inequalities provide a framework through which many optimisation problems can be solved, in particular, saddle-point problems. In this paper, we study modifications to the so-called Golden Ratio ALgorithm (GRAAL) for variational inequalities—a method which uses a fully explicit adaptive step-size and provides convergence results under local Lipschitz assumptions without requiring backtracking. We present and analyse two Bregman modifications to GRAAL: the first uses a fixed step size and converges under global Lipschitz assumptions, and the second uses an adaptive step-size rule. Numerical performance of the former method is demonstrated on a bimatrix game arising in network communication, and of the latter on two problems, namely, power allocation in Gaussian communication channels and N -person Cournot completion games. In all of these applications, an appropriately chosen Bregman distance simplifies the projection steps computed as part of the algorithm.

Keywords Variational inequality · Saddle-point problems · Bregman distance · Local Lipschitz · Adaptive step size

Mathematics Subject Classification 47J20 · 49J40 · 65K15 · 65Y20

1 Introduction

Let $\mathcal{X}, \mathcal{Y}$ be real, finite-dimensional Hilbert spaces. In this work, we consider *saddle-point problems* of the form

Communicated by Aviv Gibali.

✉ Daniel J. Uteda
duteda@student.unimelb.edu.au

Matthew K. Tam
matthew.tam@unimelb.edu.au

¹ School of Mathematics and Statistics, The University of Melbourne, Parkville, VIC 3010, Australia

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \Phi(x, y) := \psi(x) + f(x, y) - \zeta(y), \quad (1)$$

where

- $f: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ is convex–concave and continuously differentiable.
- $\psi: \mathcal{X} \rightarrow (-\infty, +\infty]$, $\zeta: \mathcal{Y} \rightarrow (-\infty, +\infty]$ are proper, lower semicontinuous (l.s.c.), and convex.

Since many non-smooth optimisation problems can be cast in the form of (1), it is a useful and heavily studied tool in itself [1, 2, 20, 25, 37, 44]. However, rather than attempt to solve (1) in its current form, it is far more convenient, even within the aforementioned references, to follow in the steps of Korpelevič [34] and Popov [47], by casting it as the *Variational Inequality* (VI):

$$\text{Find } z^* \in \mathcal{H} \text{ such that } \langle F(z^*), z - z^* \rangle + g(z) - g(z^*) \geq 0 \quad \forall z \in \mathcal{H}, \quad (2)$$

where $z = (x, y) \in \mathcal{H} := \mathcal{X} \oplus \mathcal{Y}$, and

$$F(z) = (\nabla_x f(x, y), -\nabla_y f(x, y)), \quad g(z) = \psi(x) + \zeta(y). \quad (3)$$

and the variable $z^* := (x^*, y^*)$ shown in (2) characterises the solution (x^*, y^*) to (1).

Many methods (see, for instance, [21, 43, 46]) for solving (2) require global Lipschitz continuity of the operator F . However, this assumption is often too strong to hold in practice. Even when F is globally Lipschitz continuous, knowledge of its Lipschitz constant is usually required as input to the chosen algorithm and determining this constant is typically more difficult than solving the original problem. Moreover, even if F is globally Lipschitz and its global Lipschitz constant is known, then, as the step size is inversely related to the Lipschitz constant, a constant step-size rule can be too conservative. This is particularly unnecessary if the generated sequence lies entirely within a region where a *local* Lipschitz constant is small (relative to the size of the global constant).

Therefore, it is beneficial to instead define a step-size sequence which attempts to approximate a local Lipschitz constant with respect to the point iterates. The standard approach then is to generate a step-size sequence via a *backtracking* procedure (see [11, 15, 29, 38, 39, 42] and references therein). While avoiding each of the shortcomings listed above, such methods can become expensive when considering the overall run-time of the algorithm, due to the arbitrarily large number of steps taken during the backtracking procedure within each iteration. An emerging alternative is that of *adaptive step sizes* [3, 40, 41], which accomplish the same goals as backtracking methods without the need for backtracking, i.e. the step-size update is fully explicit. In particular, the *adaptive Golden RAtio ALgorithm* (*aGRAAL*) [40] (as stated in Algorithm 1), named as such because of its relationship with the *Golden Ratio* $\varphi = \frac{1+\sqrt{5}}{2}$, solves (2) and is the method we focus on here.

One other way to potentially improve methods is to replace the Euclidean distance in the proximal operator with a non-Euclidean family of distance-like functions called the

Algorithm 1: The Adaptive Golden RAtio ALgorithm (aGRAAL) [40].**Input:** Initial points $z_0, \bar{z}_0 \in \mathcal{H}$, initial step-size $\lambda_0 > 0, \lambda_{\max} \gg 0, \phi \in (1, \varphi]$.**Initialisation:** Set $z_1 = \bar{z}_0, \theta_0 = 1, \rho = \frac{1}{\phi} + \frac{1}{\phi^2}$.**for** $k = 1, 2, \dots$ **do**

Calculate the step-size:

$$\lambda_k = \min \left\{ \rho \lambda_{k-1}, \frac{\phi \theta_{k-1}}{4 \lambda_{k-1}} \frac{\|z_k - z_{k-1}\|^2}{\|F(z_k) - F(z_{k-1})\|^2}, \lambda_{\max} \right\}. \quad (4)$$

Compute the next iterates:

$$\bar{z}_k = \frac{(\phi - 1)z_k + \bar{z}_{k-1}}{\phi} \quad (5)$$

$$z_{k+1} = \text{prox}_{\lambda_k g}(\bar{z}_k - \lambda_k F(z_k)) \quad (6)$$

 Update: $\theta_k = \frac{\lambda_k \phi}{\lambda_{k-1}}$.

Bregman distance. Such methods, for solving (2), can be found in existing literature [15, 22, 23, 26, 27, 30–32, 45, 48]. Interestingly, most of these methods require a Lipschitz assumption but do not require knowledge of the Lipschitz constant. However, these employ a backtracking procedure and/or a non-increasing step-size sequence, such as that found in [28], whereas the step size of our new method is fully explicit and is allowed to increase slightly at each iteration.

In this paper, we investigate Bregman modifications to Algorithm 1. To this end, we begin by proposing the *Bregman-Golden RAtio ALgorithm (B-GRAAL)*, a Bregman version of the fixed step-size *Golden RAtio ALgorithm (GRAAL)*, and prove convergence of our new method in full. We then present an adaptive version of B-GRAAL, or similarly, a Bregman modification to Algorithm 1, which we refer to as the *Bregman-adaptive Golden RAtio ALgorithm (B-aGRAAL)*. Although we only provide a convergence analysis of B-aGRAAL in a restrictive setting, we observe it to work numerically outside of this setting.

One advantage of our new method is the flexibility provided by the Bregman proximal operator. In the context of convex–concave games, for instance, this modified operator arises as the projection onto the probability simplex which has a simple closed-form expression with respect to the *Kullback–Leibler (KL) divergence* but not with respect to the standard Euclidean distance. In fact, the Euclidean projection requires an $O(n \log n)$ time algorithm in n -dimensions [18, 19, 50], whereas the KL projection only requires $O(n)$ time (see, for instance [10, Section 5] and [35, Section 4.4]). Another advantage of these modifications in the constrained optimisation case is that it is sometimes possible to choose a Bregman distance whose domain is the constraint set so as to make the feasibility of the iterates implicit.

The remainder of this paper is structured as follows. In Sect. 2, we collect preliminary results for use in our analysis. In Sect. 3, we present our fixed step-size method and a proof of convergence. In Sect. 4, we present our adaptive method with some partial analysis. Section 5 contains experimental results. Firstly, we compare the fixed step-size method with the Euclidean distance and the KL divergence on a matrix

game between two players. Secondly, we make the same comparison for the adaptive method on a power allocation problem in a Gaussian communication channel. Finally, we apply the adaptive method to an N -person Cournot oligopoly model with appropriately chosen Bregman distances over a closed box. We then conclude this paper by presenting some directions for further research.

2 Preliminaries

Throughout this work, $\mathcal{H}$ denotes a real, finite-dimensional Hilbert space with inner-product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \|$. Given an extended real-valued function $f: \mathcal{H} \rightarrow (-\infty, +\infty]$, its *domain* is denoted $\text{dom } f := \{x \in \mathcal{H}: f(x) < +\infty\}$. Its *subdifferential* at $x \in \text{dom } f$ is given by

$$\partial f(x) := \{v \in \mathcal{H}: f(x) - f(y) - \langle v, x - y \rangle \leq 0 \quad \forall y \in \mathcal{H}\},$$

and defined as $\partial f(x) := \emptyset$ for $x \notin \text{dom } f$. The *indicator function* of a set $K \subseteq \mathcal{H}$ is written ι_K and takes the value 0 for $x \in K$ and $+\infty$ otherwise.

A proper, l.s.c., convex function $h: \mathcal{H} \rightarrow (-\infty, +\infty]$ is called *Legendre* [12, Definition 7.1.1] if it is strictly convex on every convex subset of $\text{int dom } h := \{x \in \mathcal{H}: \partial h(x) \neq \emptyset\}$ and differentiable on $\text{int dom } h \neq \emptyset$ such that $\|\nabla h(x)\| \rightarrow \infty$ whenever x approaches the boundary of $\text{dom } h$. The *convex conjugate* of h written as $h^*: \mathcal{H} \rightarrow (-\infty, +\infty]$ is the given by

$$h^*(x^*) := \sup_{x \in \mathcal{H}} \{\langle x^*, x \rangle - h(x)\}.$$

When h is merely differentiable on $\text{int dom } h$, the *Bregman distance* generated by h is the function $D_h: \mathcal{H} \times \text{int dom } h \rightarrow (-\infty, +\infty]$ given by

$$D_h(x, y) := h(x) - h(y) - \langle \nabla h(y), x - y \rangle.$$

When h is also convex, D_h is nonnegative, and when h is σ -strongly convex, D_h satisfies

$$D_h(x, y) \geq \frac{\sigma}{2} \|x - y\|^2 \quad \forall (x, y) \in \mathcal{H} \times \text{int dom } h.$$

We begin by collecting some general properties of the Bregman distance.

Proposition 2.1 (Properties of the Bregman distance) *Let $h: \mathcal{H} \rightarrow (-\infty, +\infty]$ be Legendre. Then, the following assertions hold.*

(a) (three-point identity) *For all $x, y \in \text{int dom } h$ and $z \in \text{dom } h$, we have*

$$D_h(z, x) - D_h(z, y) - D_h(y, x) = \langle \nabla h(x) - \nabla h(y), y - z \rangle.$$

(b) For all $x, y \in \text{int dom } h$, we have

$$D_h(x, y) = D_{h^*}(\nabla h(y), \nabla h(x))$$

(c) Let $x \in \text{dom } h$, $y, u, v \in \text{int dom } h$, and $\alpha \in \mathbb{R}$. Suppose additionally that $\nabla h(y) = \alpha \nabla h(u) + (1 - \alpha) \nabla h(v)$. Then,

$$D_h(x, y) = \alpha [D_h(x, u) - D_h(y, u)] + (1 - \alpha) [D_h(x, v) - D_h(y, v)].$$

Proof (a) See, for instance, [49, Lemma 2.2] and the paragraph immediately after.

(b) See, for instance, [4, Theorem 3.7(v)]. (c) By using the definition of D_h , together with the assumption $\nabla h(y) = \alpha \nabla h(u) + (1 - \alpha) \nabla h(v)$, we obtain

$$\begin{aligned} D_h(x, y) &= h(x) - h(y) - \langle \nabla h(y), x - y \rangle \\ &= \alpha [h(x) - h(y) - \langle \nabla h(u), x - y \rangle] + (1 - \alpha) [h(x) - h(y) - \langle \nabla h(v), x - y \rangle] \\ &= \alpha [h(x) - h(u) - \langle \nabla h(u), x - u \rangle - h(y) + h(u) - \langle \nabla h(u), u - y \rangle] \\ &\quad + (1 - \alpha) [h(x) - h(v) - \langle \nabla h(v), x - v \rangle - h(y) + h(v) - \langle \nabla h(v), v - y \rangle] \\ &= \alpha [D_h(x, u) - D_h(y, u)] + (1 - \alpha) [D_h(x, v) - D_h(y, v)]. \end{aligned}$$

This completes the proof. $\square$

Remark 2.1 When $h = \frac{1}{2} \|\cdot\|^2$, the expression shown in Proposition 2.1(c) simplifies to the Euclidean identity

$$\forall x, y \in \mathcal{H}, \alpha \in \mathbb{R} \quad \|\alpha x + (1 - \alpha)y\|^2 = \alpha \|x\|^2 + (1 - \alpha) \|y\|^2 - \alpha(1 - \alpha) \|x - y\|^2.$$

We now turn our attention to operators defined in terms of the Bregman divergence. The (left) Bregman proximal operator of a function $f: \mathcal{H} \rightarrow (-\infty, +\infty]$ is the (potentially set-valued) operator given by

$$\mathbf{prox}_f^h(y) := \underset{x \in \mathcal{H}}{\operatorname{argmin}} \{f(x) + D_h(x, y)\} \quad \forall y \in \text{int dom } h. \quad (7)$$

Since we will only require the left Bregman proximal operator in this work, will omit the qualifier “left” from here onwards. For further details on the analogous “right Bregman proximal operator”, the reader is referred to [13]. The (left) Bregman projection onto C is the (left) Bregman proximal operator of ι_C , that is,

$$P_C^h(y) := \mathbf{prox}_{\iota_C}^h(y) = \underset{x \in C}{\operatorname{argmin}} D_h(x, y) \quad \forall y \in \text{int dom } h.$$

Next, we collect properties of the Bregman proximal operator for use in our subsequence algorithm analysis.

Proposition 2.2 (Bregman proximal operator) *Let $f: \mathcal{H} \rightarrow (-\infty, +\infty]$ be proper, l.s.c., convex and let $h: \mathcal{H} \rightarrow (-\infty, +\infty]$ be Legendre such that $\text{int dom } h \cap \text{dom } f \neq \emptyset$.*

- (a) $\text{range}(\text{prox}_f^h) \subseteq \text{int dom } h \cap \text{dom } f$.
- (b) prox_f^h is single-valued on $\text{dom}(\text{prox}_f^h) \subseteq \text{int dom } h$. Moreover, if $h + f$ is super-coercive, that is, $\lim_{\|x\| \rightarrow +\infty} \frac{h(x) + f(x)}{\|x\|} = +\infty$, then $\text{dom}(\text{prox}_f^h) = \text{int dom } h$.
- (c) Let $y \in \text{dom}(\text{prox}_f^h)$. Then, $x = \text{prox}_f^h(y)$ if and only if, for all $u \in \mathcal{H}$, we have

$$f(x) - f(u) \leq \langle \nabla h(y) - \nabla h(x), x - u \rangle. \quad (8)$$

- (d) Let $y, y' \in \text{dom}(\text{prox}_f^h)$, $x = \text{prox}_f^h(y)$ and $x' = \text{prox}_f^h(y')$. Then

$$0 \leq \langle \nabla h(x) - \nabla h(x'), x - x' \rangle \leq \langle \nabla h(y) - \nabla h(y'), x - x' \rangle. \quad (9)$$

Proof (a) See [7, Proposition 3.23(v)(b)], noting that the sum rule holds for f and h since $\text{int dom } h \cap \text{dom } f \neq \emptyset$ and thus $\text{dom } \partial(f+h) = \text{dom}(\partial f + \nabla h) \subseteq \text{dom } f \cap \text{int dom } h$. (b) The first part follows by combining (a) and [7, Proposition 3.22 (ii)(d)]. For the second part, see [7, Proposition 3.21(vii)]. (c) Since the sum rule holds for f and h , the first-order optimality condition implies $x = \text{prox}_f^h(y)$ if and only if $\nabla h(y) - \nabla h(x) \in \partial f(x)$. The latter is equivalent to (8). (d) The first inequality in (9) follows from convexity of h . To show the second, we apply (8) with $u = x'$ to see

$$f(x) - f(x') \leq \langle \nabla h(y) - \nabla h(x), x - x' \rangle,$$

and similarly,

$$f(x') - f(x) \leq \langle \nabla h(y') - \nabla h(x'), x' - x \rangle.$$

Then, adding these inequalities gives the desired result. $\square$

Remark 2.2 Parts (a), (b), (d) of Proposition 2.2 also apply to the Bregman resolvent [13], which is defined for a set-valued operator $A: \mathcal{H} \rightrightarrows \mathcal{H}$ as $R_A^h = (\nabla h + A)^{-1} \circ \nabla h$. To see that the resolvent generalises the proximal operator, we refer to [7, Proposition 3.22 (ii)(a)].

3 The Bregman-Golden Ratio Algorithm

In this section, we consider the Variational Inequality (VI) problem

$$\langle F(z^*), z - z^* \rangle + g(z) - g(z^*) \geq 0 \quad \forall z \in \mathcal{H}, \quad (10)$$

where we assume that

A.1 $g: \mathcal{H} \rightarrow (-\infty, +\infty]$ is proper, l.s.c., convex.

- A.2** $h: \mathcal{H} \rightarrow (-\infty, +\infty]$ is continuously differentiable, Legendre, and σ -strongly convex. In addition, we will also require that $D_h(x, x_n) \rightarrow 0$ for every sequence $(x_n) \subseteq \text{int dom } h$ that converges to some $x \in \text{dom } h$.
- A.3** $F: \mathcal{H} \rightarrow \mathcal{H}$ is monotone over $\text{dom } g \cap \text{int dom } h \neq \emptyset$.
- A.4** $\Omega := S \cap \text{dom } h \neq \emptyset$ where S denotes the solution set of (10).

Remark 3.1 Assumption **A.2** is common in the literature concerning Bregman first-order methods [8, 9, 17, 49]. In particular, the limit condition holds when ∇h is continuous and $\text{dom } h$ is open. Indeed, in this case, σ -strong convexity of h implies that h^* is $\frac{1}{\sigma}$ -smooth [33, Theorem 6], and so applying Proposition 2.1(b) gives

$$D_h(x, x_n) = D_{h^*}(\nabla h(x_n), \nabla h(x)) \leq \frac{1}{2\sigma} \|\nabla h(x_n) - \nabla h(x)\|^2 \rightarrow 0.$$

The significance of $\text{dom } h$ being open here is that h is differentiable at $x \in \text{dom } h$; however, we observe that the same condition can still hold if $\text{dom } h$ is closed. In particular, it also holds for the KL divergence (see, for instance, [17, Example 2.1]).

Our proposed algorithm for solving (10) when F is L -Lipschitz is called the *Bregman-Golden RAtio ALgorithm (B-GRAAL)* and is stated in Algorithm 2. Recall that $\varphi := \frac{1+\sqrt{5}}{2}$ denotes the *Golden Ratio*, which satisfies $\varphi^2 = \varphi + 1$.

Algorithm 2: The Bregman-Golden RAtio ALgorithm (B-GRAAL)

Input: Initial points $z_1, \bar{z}_0 \in \text{int dom } h$ and a step-size $\lambda \in (0, \frac{\sigma\varphi}{2L}]$.

for $k = 1, 2, \dots$ **do**

 Compute the next iterates:

$$\bar{z}_k = (\nabla h)^{-1} \left(\frac{(\varphi - 1)\nabla h(z_k) + \nabla h(\bar{z}_{k-1})}{\varphi} \right) \quad (11)$$

$$z_{k+1} = \underset{z \in \mathcal{H}}{\text{argmin}} \left\{ \langle F(z_k), z - z_k \rangle + g(z) + \frac{1}{\lambda} D_h(z, \bar{z}_k) \right\}. \quad (12)$$

The following lemma establishes the well-definedness of the sequences generated by the Bregman-GRAAL.

Lemma 3.1 Suppose Assumptions **A.1–A.2** hold. Then, the sequences $(\bar{z}_k)$ and (z_k) generated by Algorithm 2 are well defined. Moreover, $(\bar{z}_k) \subseteq \text{int dom } h$ and $(z_k) \subseteq \text{int dom } h \cap \text{dom } g$.

Proof Suppose by way of induction that $z_k, \bar{z}_{k-1} \in \text{int dom } h$ for some $k \geq 1$. Since the gradient $\nabla h: \text{int dom } h \rightarrow \text{int}(\text{dom } h^*)$ is a bijection [12, Theorem 7.3.7], it follows that $\nabla h(z_k), \nabla h(\bar{z}_{k-1}) \in \text{int}(\text{dom } h^*)$. As $\text{int}(\text{dom } h^*)$ is a convex set, (11) implies that $\nabla h(\bar{z}_k) \in \text{int}(\text{dom } h^*)$ which establishes that $\bar{z}_k \in \text{int dom } h$. Next, we observe that $z_{k+1} = \text{prox}_{\lambda f}^h(\bar{z}_k)$, where $f(z) := \langle F(z_k), z - z_k \rangle + g(z)$. Note that $\text{dom } f = \text{dom } g$. Since $\lambda f + h$ is σ -strongly convex, it is supercoercive by [5,

Corollary 11.16] and so Proposition 2.2(a)-(b) shows that $\text{prox}_{\lambda f}^h$ is single-valued with $\text{range}(\text{prox}_{\lambda f}^h) \subseteq \text{int dom } h \cap \text{dom } g$ and therefore $z_{k+1} \in \text{int dom } h \cap \text{dom } g$. $\square$

Remark 3.2 The Bregman proximal step shown in (12) can be expressed in terms of the Bregman proximal operator: $z_{k+1} = \text{prox}_{\lambda f}^h(\bar{z}_k)$, where $f(z) = \langle F(z_k), z - z_k \rangle + g(z)$. Equivalently, $z_{k+1} = \text{prox}_{\lambda g}^h((\nabla h)^{-1}(\nabla h(\bar{z}_k) - \lambda F(z_k)))$, due to the first-order optimality condition in Proposition 2.2(c).

The following lemma is key in our convergence analysis.

Lemma 3.2 Suppose Assumptions A.1–A.4 hold and that F is L -Lipschitz continuous on $\text{dom } g \cap \text{int dom } h$. Let $z^* \in \Omega$ be arbitrary. Then, the sequences (z_k) , $(\bar{z}_k)$ generated by Algorithm 2 satisfy

$$\begin{aligned} 0 &\leq (\varphi + 1)D_h(z^*, \bar{z}_{k+1}) + \frac{\varphi}{2}D_h(z_{k+1}, z_k) - \varphi D_h(z_{k+1}, \bar{z}_{k+1}) \\ &\leq (\varphi + 1)D_h(z^*, \bar{z}_k) + \frac{\varphi}{2}D_h(z_k, z_{k-1}) - \varphi D_h(z_k, \bar{z}_k) \\ &\quad - \left(1 - \frac{1}{\varphi}\right)D_h(z_{k+1}, \bar{z}_k). \end{aligned} \quad (13)$$

Proof By first applying Proposition 2.2(c) with $f(z) := \lambda(\langle F(z_k), z - z_k \rangle + g(z))$, $u := z \in \text{dom } h \cap \text{dom } g$ arbitrary, $x := z_{k+1}$ and $y := \bar{z}_k$, followed by the three-point identity (Proposition 2.1(a)) we obtain

$$\begin{aligned} &\lambda(\langle F(z_k), z_{k+1} - z \rangle + g(z_{k+1}) - g(z)) \\ &\leq \langle \nabla h(\bar{z}_k) - \nabla h(z_{k+1}), z_{k+1} - z \rangle \\ &= D_h(z, \bar{z}_k) - D_h(z, z_{k+1}) - D_h(z_{k+1}, \bar{z}_k). \end{aligned} \quad (14)$$

Shifting the index in (14) (by setting $k \equiv k - 1$), setting $z := z_{k+1}$, and using $\varphi \nabla h(\bar{z}_k) = (\varphi - 1) \nabla h(z_k) + \nabla h(\bar{z}_{k-1})$ followed by the three-point identity (Proposition 2.1(a)) gives

$$\begin{aligned} &\lambda(\langle F(z_{k-1}), z_k - z_{k+1} \rangle + g(z_k) - g(z_{k+1})) \\ &\leq \langle \nabla h(\bar{z}_{k-1}) - \nabla h(z_k), z_k - z_{k+1} \rangle \\ &= \varphi \langle \nabla h(\bar{z}_k) - \nabla h(z_k), z_k - z_{k+1} \rangle \\ &= \varphi [D_h(z_{k+1}, \bar{z}_k) - D_h(z_{k+1}, z_k) - D_h(z_k, \bar{z}_k)]. \end{aligned} \quad (15)$$

Let $z^* \in \Omega$. Setting $z := z^*$ in (14), summing with (15) and rearranging yields

$$\begin{aligned} &\lambda(\langle F(z_k), z_k - z^* \rangle + g(z_k) - g(z^*)) \\ &\leq D_h(z^*, \bar{z}_k) - D_h(z^*, z_{k+1}) - D_h(z_{k+1}, \bar{z}_k) \\ &\quad + \varphi [D_h(z_{k+1}, \bar{z}_k) - D_h(z_{k+1}, z_k) - D_h(z_k, \bar{z}_k)] \\ &\quad + \lambda \langle F(z_k) - F(z_{k-1}), z_k - z_{k+1} \rangle. \end{aligned} \quad (16)$$

We observe that the left side of (16) is nonnegative as a consequence of (10) and A.3:

$$0 \leq \langle F(z^*), z_k - z^* \rangle + g(z_k) - g(z^*) \leq \langle F(z_k), z_k - z^* \rangle + g(z_k) - g(z^*). \quad (17)$$

To estimate the final term in (16), we use the Cauchy–Schwarz inequality, L -Lipschitz continuity of F , σ -strong convexity of h and the inequality $\lambda \leq \frac{\sigma\varphi}{2L}$ to obtain

$$\begin{aligned} \lambda \langle F(z_k) - F(z_{k-1}), z_k - z_{k+1} \rangle &\leq \lambda L \|z_k - z_{k-1}\| \|z_k - z_{k+1}\| \\ &\leq \frac{\lambda L}{2} \left(\|z_k - z_{k-1}\|^2 + \|z_k - z_{k+1}\|^2 \right) \\ &\leq \frac{\varphi}{2} (D_h(z_k, z_{k-1}) + D_h(z_{k+1}, z_k)). \end{aligned} \quad (18)$$

Combining (16), (17) and (18) gives

$$\begin{aligned} D_h(z^*, z_{k+1}) &\leq D_h(z^*, \bar{z}_k) + (\varphi - 1)D_h(z_{k+1}, \bar{z}_k) - \varphi D_h(z_k, \bar{z}_k) \\ &\quad - \frac{\varphi}{2} D_h(z_{k+1}, z_k) + \frac{\varphi}{2} D_h(z_k, z_{k-1}). \end{aligned} \quad (19)$$

Now, applying Proposition 2.1(c) with

$$\nabla h(z_{k+1}) = \frac{\varphi \nabla h(\bar{z}_{k+1}) - \nabla h(\bar{z}_k)}{\varphi - 1} = (\varphi + 1) \nabla h(\bar{z}_{k+1}) - \varphi \nabla h(\bar{z}_k)$$

and rearranging yields

$$\begin{aligned} (\varphi + 1)D_h(z^*, \bar{z}_{k+1}) &= D_h(z^*, z_{k+1}) + (\varphi + 1)D_h(z_{k+1}, \bar{z}_{k+1}) \\ &\quad + \varphi [D_h(z^*, \bar{z}_k) - D_h(z_{k+1}, \bar{z}_k)]. \end{aligned} \quad (20)$$

Combining (19) and (20), followed by collecting like-terms, gives

$$\begin{aligned} (\varphi + 1)D_h(z^*, \bar{z}_{k+1}) &+ \frac{\varphi}{2} D_h(z_{k+1}, z_k) - \varphi D_h(z_{k+1}, \bar{z}_{k+1}) \\ &\leq (\varphi + 1)D_h(z^*, \bar{z}_k) + \frac{\varphi}{2} D_h(z_k, z_{k-1}) - \varphi D_h(z_k, \bar{z}_k) \\ &\quad + D_h(z_{k+1}, \bar{z}_{k+1}) - D_h(z_{k+1}, \bar{z}_k). \end{aligned} \quad (21)$$

Since $\nabla h(\bar{z}_{k+1}) = \frac{\varphi-1}{\varphi} \nabla h(z_{k+1}) + \frac{1}{\varphi} \nabla h(\bar{z}_k)$, Proposition 2.1(c) gives

$$\begin{aligned} D_h(z_{k+1}, \bar{z}_{k+1}) &= \frac{\varphi - 1}{\varphi} [D_h(z_{k+1}, z_{k+1}) - D_h(\bar{z}_{k+1}, z_{k+1})] \\ &\quad + \frac{1}{\varphi} [D_h(z_{k+1}, \bar{z}_k) - D_h(\bar{z}_{k+1}, \bar{z}_k)] \\ &\leq \frac{1}{\varphi} D_h(z_{k+1}, \bar{z}_k). \end{aligned} \quad (22)$$

Combining (21) and (22) establishes the second inequality in (13). To show the first inequality in (13), we apply the three-point identity (Proposition 2.1(a)) to see that

$$\begin{aligned}
 D_h(z^*, z_{k+1}) + D_h(z_{k+1}, \bar{z}_{k+1}) &= D_h(z^*, \bar{z}_{k+1}) + \langle \nabla h(\bar{z}_{k+1}) - \nabla h(z_{k+1}), z^* - z_{k+1} \rangle \\
 &= D_h(z^*, \bar{z}_{k+1}) + \frac{1}{\varphi} \langle \nabla h(\bar{z}_k) - \nabla h(z_{k+1}), z^* - z_{k+1} \rangle. \\
 &\leq D_h(z^*, \bar{z}_{k+1}) + \frac{\lambda}{\varphi} \left(\langle F(z_k), z^* - z_{k+1} \rangle - g(z_{k+1}) + g(z^*) \right) \\
 &= D_h(z^*, \bar{z}_{k+1}) + \frac{\lambda}{\varphi} \langle F(z_k) - F(z_{k+1}), z^* - z_{k+1} \rangle \\
 &\quad + \frac{\lambda}{\varphi} \left(\langle F(z_{k+1}), z^* - z_{k+1} \rangle - g(z_{k+1}) + g(z^*) \right) \\
 &\leq D_h(z^*, \bar{z}_{k+1}) + \frac{\lambda}{\varphi} \langle F(z_k) - F(z_{k+1}), z^* - z_{k+1} \rangle.
 \end{aligned} \tag{23}$$

Using L -Lipschitz continuity of F and σ -strong convexity of h gives

$$\frac{\lambda}{\varphi} \langle F(z_k) - F(z_{k+1}), z^* - z_{k+1} \rangle \leq \frac{1}{2} D_h(z_{k+1}, z_k) + \frac{1}{2} D_h(z^*, z_{k+1}). \tag{24}$$

On substituting (24) back into (23) and rearranging, we obtain

$$\begin{aligned}
 D_h(z_{k+1}, \bar{z}_{k+1}) &\leq D_h(z^*, \bar{z}_{k+1}) + \frac{1}{2} D_h(z_{k+1}, z_k) - \frac{1}{2} D_h(z^*, z_{k+1}) \\
 &\leq D_h(z^*, \bar{z}_{k+1}) + \frac{1}{2} D_h(z_{k+1}, z_k),
 \end{aligned} \tag{25}$$

and therefore

$$\begin{aligned}
 0 &\leq D_h(z^*, \bar{z}_{k+1}) + \frac{1}{2} D_h(z_{k+1}, z_k) - D_h(z_{k+1}, \bar{z}_{k+1}) \\
 \implies 0 &\leq (\varphi + 1) D_h(z^*, \bar{z}_{k+1}) + \frac{\varphi}{2} D_h(z_{k+1}, z_k) - \varphi D_h(z_{k+1}, \bar{z}_{k+1}),
 \end{aligned}$$

which establishes the first inequality of (13) and thus completes the proof. $\square$

The following is our main result regarding convergence of the Bregman GRAAL with fixed step size.

Theorem 3.1 *Suppose Assumptions A.1–A.4 hold and that F is L -Lipschitz continuous on $\text{dom } g \cap \text{int dom } h$. Then, the sequences (z_k) and $(\bar{z}_k)$ generated by Algorithm 2 converge to a point in Ω .*

Proof Let $z^* \in \Omega$ be arbitrary, and let (η_k) denote the sequence given by

$$\eta_k := (\varphi + 1) D_h(z^*, \bar{z}_k) + \frac{\varphi}{2} D_h(z_k, z_{k-1}) - \varphi D_h(z_k, \bar{z}_k) \quad \forall k \in \mathbb{N}. \tag{26}$$

Lemma 3.2 implies that $\lim_{k \rightarrow \infty} \eta_k$ exists and $D_h(z_{k+1}, \bar{z}_k) \rightarrow 0$ as $k \rightarrow \infty$. Referring to (22), it follows that $D_h(z_{k+1}, \bar{z}_{k+1}) \rightarrow 0$. Also, by applying Proposition 2.1(c) with the identity $\nabla h(z_k) = (\varphi + 1)\nabla h(\bar{z}_k) - \varphi\nabla h(\bar{z}_{k-1})$, we obtain

$$\begin{aligned} D_h(z_{k+1}, z_k) &= (\varphi + 1)[D_h(z_{k+1}, \bar{z}_k) - D_h(z_k, \bar{z}_k)] \\ &\quad - \varphi[D_h(z_{k+1}, \bar{z}_{k-1}) - D_h(z_k, \bar{z}_{k-1})] \\ &\leq (\varphi + 1)D_h(z_{k+1}, \bar{z}_k) + \varphi D_h(z_k, \bar{z}_{k-1}) \rightarrow 0. \end{aligned}$$

Altogether, we have that

$$\lim_{k \rightarrow \infty} \eta_k = (\varphi + 1) \lim_{k \rightarrow \infty} D_h(z^*, \bar{z}_k),$$

and, in particular, $\lim_{k \rightarrow \infty} D_h(z^*, \bar{z}_k)$ exists.

Next, using σ -strong convexity of h , we deduce that $z_{k+1} - \bar{z}_k \rightarrow 0$, and that (z_k) and $(\bar{z}_k)$ are bounded. Thus, let $\bar{z} \in \mathcal{H}$ be a cluster point of $(\bar{z}_k)$. Then, there exists a subsequence $(\bar{z}_{k_j})$ such that $\bar{z}_{k_j} \rightarrow \bar{z}$ and $z_{k_j+1} \rightarrow \bar{z}$ as $j \rightarrow \infty$. Now, recalling (14) gives

$$\begin{aligned} &\lambda \left(\langle F(z_{k_j}), z_{k_j+1} - z \rangle + g(z_{k_j+1}) - g(z) \right) \\ &\leq \langle \nabla h(\bar{z}_{k_j}) - \nabla h(z_{k_j+1}), z_{k_j+1} - z \rangle \quad \forall z \in \mathcal{H}, \end{aligned}$$

and taking the limit-infimum of both sides as $j \rightarrow \infty$ shows that $\bar{z} \in \Omega$. Since $z^* \in \Omega$ was chosen in Lemma 3.2 to be arbitrary, we can now set $z^* = \bar{z}$. It then follows that $\lim_{j \rightarrow \infty} D_h(z^*, \bar{z}_{k_j}) = 0$, and consequently, $\lim_{j \rightarrow \infty} \eta_{k_j} = 0$. Also note that for $n \geq k_j$, we have $\eta_n \leq \eta_{k_j}$ from Lemma 3.2, and therefore,

$$(\varphi + 1) \lim_{n \rightarrow \infty} D_h(z^*, \bar{z}_n) = \lim_{n \rightarrow \infty} \eta_n \leq \lim_{j \rightarrow \infty} \eta_{k_j} = 0,$$

and therefore $\bar{z}_k \rightarrow z^*$ from strong convexity. The fact that $z_k \rightarrow z^*$ follows since $z_k - \bar{z}_k \rightarrow 0$. $\square$

Remark 3.3 In the special case where $h = \|\cdot\|^2$, Algorithm 2 recovers the Euclidean GRAAL with fixed step size from [40, Section 2] and the conclusions of Theorem 3.1 recover [40, Theorem 1]. Despite this, the proof provided here is new and not the same as the one in [40, Theorem 1] even when specialised to the Euclidean case. Indeed, [40, Theorem 1] proceeds by establishing the inequality

$$\begin{aligned} &(\varphi + 1)\|\bar{z}_{k+1} - z^*\|^2 + \frac{\varphi}{2}\|z_{k+1} - z_k\|^2 \\ &\leq (\varphi + 1)\|\bar{z}_k - z^*\|^2 + \frac{\varphi}{2}\|z_k - z_{k-1}\|^2 - \varphi\|z_k - \bar{z}_k\|^2, \end{aligned} \quad (27)$$

which is different to Lemma 3.2. Interestingly, (27) can be deduced from (21) in Lemma 3.2 by using the equality

$$\|z_{k+1} - \bar{z}_{k+1}\|^2 = \frac{1}{\varphi^2} \|z_{k+1} - \bar{z}_k\|^2,$$

which applies in the Euclidean case, in place of (22) followed by the identity $\varphi^2 = \varphi + 1$. Note also that the inequality (22) is already weaker than the inequality $\|z_{k+1} - \bar{z}_{k+1}\|^2 \leq \frac{1}{\varphi^2} \|z_{k+1} - \bar{z}_k\|^2$.

3.1 Linear Convergence of B-GRAAL

In this subsection, we investigate linear convergence of Algorithm 2. Recall that a sequence $(z_k) \subset \mathcal{H}$ is said to converge *Q-linearly* to $z \in \mathcal{H}$ if there exists some $q \in (0, 1)$ such that $\|z_{k+1} - z\| < q\|z_k - z\|$ for all k sufficiently large. A sequence (z_k) is said to converge *R-linearly* if $\|z_k - z\| \leq \gamma_k$ for sufficiently large k and a sequence $(\gamma_k) \subseteq \mathbb{R}$ converging *Q-linearly* to 0.

In [40, Section 2.3], the sequences generated by Algorithm 1 were shown to converge *R-linearly* under the following error bound condition: there exists $\alpha, \beta > 0$ such that

$$\|z - \text{prox}_{\lambda g}(z - \lambda F(z))\| \leq \alpha \implies \text{dist}(z, S) \leq \beta \|z - \text{prox}_{\lambda g}(z - \lambda F(z))\|$$

for all z . However, the proof in [40, Section 2.3] relies heavily on the (Euclidean) triangle inequality and thus does not generalise directly to the Bregman setting. Instead, we shall assume that, for some $\mu > 0$, F satisfies

$$\langle F(x) - F(y), x - y \rangle \geq \mu D_h(x, y) \quad \forall (x, y) \in \text{dom } h \times \text{int dom } h. \quad (28)$$

When $h = \frac{1}{2} \|\cdot\|^2$, this property reduces to the standard definition of μ -strong monotonicity. When the operator F is continuous, condition (28) is weaker than the notion of μ -strongly monotone relative to h introduced in [24]. The latter has been used to show linear convergence of the Bregman proximal point algorithm [24, Theorem 3.3].

The following is our main result concerning linear convergence of Algorithm 2 under condition (28).

Theorem 3.2 *Suppose Assumptions A.1–A.4 hold, and that F is L -Lipschitz continuous on $\text{dom } g \cap \text{int dom } h$. Additionally, suppose that F satisfies condition (28) holds for some $\mu > 0$, and that $\lambda < \frac{\sigma\varphi}{2L}$. Then, the sequences (z_k) and $(\bar{z}_k)$ generated by Algorithm 2 converge *R-linearly* to the unique point in Ω .*

Proof By decreasing the value of $\mu > 0$ if necessary, we assume without loss of generality that

$$\mu < \frac{1}{2} \quad \text{and} \quad \lambda \leq \frac{\sigma\varphi\sqrt{1-\mu}}{2L}.$$

Let $z^* \in \Omega$. Since F satisfies (28), we have

$$\begin{aligned} 0 &\leq \langle F(z^*), z_k - z^* \rangle + g(z_k) - g(z^*) \\ &\leq \langle F(z_k), z_k - z^* \rangle + g(z_k) - g(z^*) - \mu D_h(z^*, z_k). \end{aligned} \quad (29)$$

Using the Cauchy–Schwarz inequality, L -Lipschitz continuity of F , σ -strong convexity of h and the inequality $\leq \frac{\sigma\varphi\sqrt{1-\mu}}{2L}$ yields

$$\begin{aligned} &\lambda \langle F(z_k) - F(z_{k-1}), z_k - z_{k+1} \rangle \\ &\leq \frac{\sigma\varphi}{2} \left(\sqrt{1-\mu} \|z_k - z_{k-1}\| \right) \cdot \|z_k - z_{k+1}\| \\ &\leq \frac{\sigma\varphi}{2} \left(\frac{(1-\mu)}{2} \|z_k - z_{k-1}\|^2 + \frac{1}{2} \|z_k - z_{k+1}\|^2 \right) \\ &\leq \frac{(1-\mu)\varphi}{2} D_h(z_k, z_{k-1}) + \frac{\varphi}{2} D_h(z_{k+1}, z_k). \end{aligned} \quad (30)$$

By following the same steps as the proof of Lemma 3.2, but (29) in place of (17) and (30) in place of (18), the analogue of (19) becomes

$$\begin{aligned} D_h(z^*, z_{k+1}) &\leq D_h(z^*, \bar{z}_k) + (\varphi - 1) D_h(z_{k+1}, \bar{z}_k) - \varphi D_h(z_k, \bar{z}_k) - \mu D_h(z^*, z_k) \\ &\quad - \frac{\varphi}{2} D_h(z_{k+1}, z_k) + \frac{(1-\mu)\varphi}{2} D_h(z_k, z_{k-1}). \end{aligned} \quad (31)$$

Applying Proposition 2.1(c) with $\nabla h(z_{k+1}) = (\varphi + 1)\nabla h(\bar{z}_{k+1}) - \varphi\nabla h(\bar{z}_k)$ gives

$$\begin{aligned} D_h(z^*, z_{k+1}) &= (\varphi + 1) [D_h(z^*, \bar{z}_{k+1}) - D_h(z_{k+1}, \bar{z}_{k+1})] \\ &\quad - \varphi [D_h(z^*, \bar{z}_k) - D_h(z_{k+1}, \bar{z}_k)]. \end{aligned} \quad (32)$$

Let (η_k) denote the sequence given by (26). Then, using the identity (32) to eliminate $D_h(z^*, z_{k+1})$ and $D_h(z^*, z_k)$ from (31) gives

$$\begin{aligned} \eta'_{k+1} &:= (\varphi + 1) D_h(z^*, \bar{z}_{k+1}) + \frac{\varphi}{2} D_h(z_{k+1}, z_k) - D_h(z_{k+1}, \bar{z}_{k+1}) \\ &\leq (1-\mu)(\varphi + 1) D_h(z^*, \bar{z}_k) + \frac{(1-\mu)\varphi}{2} D_h(z_k, z_{k-1}) + (\mu(1+\varphi) - \varphi) D_h(z_k, \bar{z}_k) \\ &\quad + \mu\varphi D_h(z^*, \bar{z}_{k-1}) - \mu\varphi D_h(z_k, \bar{z}_{k-1}) + \varphi D_h(z_{k+1}, \bar{z}_{k+1}) - D_h(z_{k+1}, \bar{z}_k) \\ &\leq (1-\mu)\eta_k + \mu\varphi D_h(z^*, \bar{z}_{k-1}), \end{aligned}$$

where we note that the final inequality uses $D_h(z_k, \bar{z}_k) \leq \frac{1}{\varphi} D_h(z_k, \bar{z}_{k-1})$. Since (η_k) is non-increasing by Lemma 3.2 and $\frac{(1-\mu)\varphi}{1-\mu(2-\mu)} \geq 1$, it follows that

$$\begin{aligned} \eta'_{k+1} &\leq (1-\mu)\eta_{k-1} + \mu(\varphi - 1)(\varphi + 1) D_h(z^*, \bar{z}_{k-1}) \\ &= (1 - \mu(2 - \varphi))(\varphi + 1) D_h(z^*, \bar{z}_{k-1}) + (1-\mu)\frac{\varphi}{2} D_h(z_{k-1}, z_{k-2}) \\ &\quad - (1-\mu)\varphi D_h(z_{k-1}, \bar{z}_{k-1}) \end{aligned}$$

$$\begin{aligned} &\leq (1 - \mu(2 - \varphi))[(\varphi + 1)D_h(z^*, \bar{z}_{k-1}) + \frac{\varphi}{2}D_h(z_{k-1}, z_{k-2}) - D_h(z_{k-1}, \bar{z}_{k-1})] \\ &= q\eta'_{k-1} \text{ where } q := (1 - \mu(2 - \varphi)) \in (0, 1). \end{aligned}$$

Thus, we have established R -linear convergence of (η'_k) to zero. Using (25) and σ -strong convexity of h , we then deduce that

$$\begin{aligned} \eta'_{k+1} &= \varphi D_h(z^*, \bar{z}_{k+1}) + \frac{\varphi - 1}{2} D_h(z_{k+1}, z_k) \\ &\quad + \left[D_h(z^*, \bar{z}_{k+1}) + \frac{1}{2} D_h(z_{k+1}, z_k) - D_h(z_{k+1}, \bar{z}_{k+1}) \right] \\ &\geq \frac{\sigma\varphi}{2} \|z^* - \bar{z}_{k+1}\|^2 + \frac{\sigma(\varphi - 1)}{4} \|z_{k+1} - z_k\|^2. \end{aligned}$$

From this, it follows that $(\bar{z}_k)$ converges R -linearly to z^* and $\|z_{k+1} - z_k\|$ converges R -linearly to 0. The latter then implies that (z_k) converges R -linearly to z^* . Since z^* was chosen arbitrarily from Ω , it must be unique. $\square$

4 The Adaptive Bregman-Golden Ratio Algorithm

In this section, we present an adaptive modification to Algorithm 2 and analyse its convergence. As with the Euclidean aGRAAL, our Bregman adaptive modification has a fully explicit step-size rule. It is presented in Algorithm 3.

Algorithm 3: The Bregman adaptive Golden Ratio ALgorithm (B-aGRAAL)

Input: Initial points $z_0, \bar{z}_0 \in \text{dom } g$, initial step-size $\lambda_0 > 0$, $\lambda_{\max} \gg 0$, $\phi \in (1, \varphi]$

Initialisation: Set $z_1 = \bar{z}_0$, $\theta_0 = 1$, choose $\rho \in \left[1, \frac{1}{\phi} + \frac{1}{\phi^2}\right]$

for $k = 1, 2, \dots$ **do**

 Calculate the step-size:

$$\lambda_k = \min \left\{ \rho\lambda_{k-1}, \frac{\sigma\phi\theta_{k-1}}{4\lambda_{k-1}} \frac{\|z_k - z_{k-1}\|^2}{\|F(z_k) - F(z_{k-1})\|^2}, \lambda_{\max} \right\}. \quad (33)$$

 Compute the next iterates:

$$\bar{z}_k = (\nabla h)^{-1} \left(\frac{(\phi - 1)\nabla h(z_k) + \nabla h(\bar{z}_{k-1})}{\phi} \right) \quad (34)$$

$$z_{k+1} = \underset{z \in \mathcal{H}}{\operatorname{argmin}} \left\{ \langle F(z_k), z - z_k \rangle + g(z) + \frac{1}{\lambda_k} D_h(z, \bar{z}_k) \right\} \quad (35)$$

 Update: $\theta_k = \frac{\lambda_k\phi}{\lambda_{k-1}}$.

We must note that in the case of $F(z_k) = F(z_{k-1})$ in (33), we adopt the convention $\frac{x}{0} = +\infty$ for all $x \in \mathbb{R}$, and therefore, $\lambda_k = \min\{\rho\lambda_{k-1}, +\infty, \lambda_{\max}\} = \min\{\rho\lambda_{k-1}, \lambda_{\max}\}$.

Observe that the step-size sequence (λ_k) in Algorithm 3 approximates the inverse of a local Lipschitz constant in the following sense:

$$\begin{aligned}\lambda_k &\leq \frac{\sigma\phi\theta_{k-1}}{4\lambda_{k-1}} \frac{\|z_k - z_{k-1}\|^2}{\|F(z_k) - F(z_{k-1})\|^2} = \frac{\sigma\theta_k\theta_{k-1}}{4\lambda_k} \frac{\|z_k - z_{k-1}\|^2}{\|F(z_k) - F(z_{k-1})\|^2} \\ &\implies \lambda_k \|F(z_k) - F(z_{k-1})\| \leq \frac{\sqrt{\sigma\theta_k\theta_{k-1}}}{2} \|z_k - z_{k-1}\|.\end{aligned}\quad (36)$$

Before giving our main convergence result for Algorithm 3, we require some preparatory lemmas. The first two are concerned with well-definedness of the algorithm and boundedness of the step-size sequence. In particular, the proof of the latter is similar to that of [40, Lemma 2], the only difference here is that we account for the updated step-size rule, and we derive explicit bounds which will become useful later.

Lemma 4.1 *Suppose Assumptions A.1–A.2 hold. Then, the sequences $(\bar{z}_k)$ and (z_k) generated by Algorithm 3 are well defined. Moreover, $(\bar{z}_k) \subseteq \text{int dom } h$ and $(z_k) \subseteq \text{int dom } h \cap \text{dom } g$.*

Proof Follows by an analogous argument to that of Lemma 3.1 but with λ replaced by λ_k and φ replaced by ϕ . $\square$

Lemma 4.2 *If (z_k) generated by Algorithm 3 is bounded and F is locally Lipschitz, then both (λ_k) and (θ_k) are bounded and separated from 0. In fact, there exists some $L > 0$ satisfying $\|F(z_k) - F(z_{k-1})\| \leq L\|z_k - z_{k-1}\|$ for all $k \in \mathbb{N}$, such that*

$$\lambda_k \geq \frac{\sigma\phi^2}{4L^2\lambda_{\max}} \quad \text{and} \quad \theta_k \geq \frac{\sigma\phi^3}{4L^2\lambda_{\max}^2} \quad \forall k \in \mathbb{N}.$$

Proof First we note that $\lambda_k \leq \lambda_{\max}$ by definition, and that $\theta_k \leq \rho\phi \leq 1 + \frac{1}{\phi}$. Since (z_k) is bounded and F is locally Lipschitz continuous, there exists $L > 0$ such that $\|F(z_k) - F(z_{k-1})\| \leq L\|z_k - z_{k-1}\|$ for all $k \in \mathbb{N}$. Without loss of generality, choose L sufficiently large so that $\lambda_i \geq \frac{\sigma\phi^2}{4L^2\lambda_{\max}}$ for $i = 0, 1$. Now, by way of induction, suppose that $\lambda_i \geq \frac{\sigma\phi^2}{4L^2\lambda_{\max}}$ for all $i = 0, \dots, k-1$. Then, there are three cases:

$$\begin{aligned}\lambda_k &= \lambda_{\max} \geq \lambda_{k-1} \geq \frac{\sigma\phi^2}{4L^2\lambda_{\max}}, \\ \lambda_k &= \rho\lambda_{k-1} > \lambda_{k-1} \geq \frac{\sigma\phi^2}{4L^2\lambda_{\max}}, \quad \text{or,} \\ \lambda_k &= \frac{\sigma\phi\theta_{k-1}}{4\lambda_{k-1}} \frac{\|z_k - z_{k-1}\|^2}{\|F(z_k) - F(z_{k-1})\|^2} \geq \frac{\sigma\phi^2}{4L^2\lambda_{k-1}} \geq \frac{\sigma\phi^2}{4L^2\lambda_{\max}}.\end{aligned}$$

In each case, the desired lower bound holds, and the bound $\theta_k \geq \frac{\sigma\phi^3}{4L^2\lambda_{\max}^2}$ follows immediately. $\square$

Our next result establishes an inequality similar, but not completely analogous, to its fixed step-size counterpart in Lemma 3.2.

Lemma 4.3 *Suppose Assumptions A.1–A.4 hold. Let $z^* \in \Omega$ be arbitrary. Then, the sequences (z_k) , $(\bar{z}_k)$ generated by Algorithm 3 satisfy.*

$$\begin{aligned} & \frac{\phi}{\phi-1} D_h(z^*, \bar{z}_{k+1}) + \frac{\theta_k}{2} D_h(z_{k+1}, z_k) - D_h(z_{k+1}, \bar{z}_{k+1}) \\ & \leq \frac{\phi}{\phi-1} D_h(z^*, \bar{z}_k) + \frac{\theta_{k-1}}{2} D_h(z_k, z_{k-1}) - \theta_k D_h(z_k, \bar{z}_k) \\ & \quad + \left(\theta_k - 1 - \frac{1}{\phi} \right) D_h(z_{k+1}, \bar{z}_k). \end{aligned} \quad (37)$$

Proof We proceed in a similar fashion as in the proof to Lemma 3.2. By first applying Proposition 2.2(c) with $f(z) := \lambda_k(\langle F(z_k), z - z_k \rangle + g(z))$, $u := z \in \text{dom } h \cap \text{dom } g$ arbitrary, $x := z_{k+1}$ and $y_0 := \bar{z}_k$, followed by the three-point identity (Proposition 2.1(a)) we obtain

$$\begin{aligned} & \lambda_k(\langle F(z_k), z_{k+1} - z \rangle + g(z_{k+1}) - g(z)) \\ & \leq \langle \nabla h(\bar{z}_k) - \nabla h(z_{k+1}), z_{k+1} - z \rangle \\ & = D_h(z, \bar{z}_k) - D_h(z, z_{k+1}) - D_h(z_{k+1}, \bar{z}_k). \end{aligned} \quad (38)$$

Shifting the index (38) (by setting $k \equiv k-1$), setting $z := z_{k+1}$ and using the fact that $\phi \nabla h(\bar{z}_k) = (\phi-1) \nabla h(z_k) + \nabla h(\bar{z}_{k-1})$ followed by the three-point identity (Proposition 2.1(a)) gives

$$\begin{aligned} & \lambda_{k-1}(\langle F(z_{k-1}), z_k - z_{k+1} \rangle + g(z_k) - g(z_{k+1})) \\ & \leq \langle \nabla h(\bar{z}_{k-1}) - \nabla h(z_k), z_k - z_{k+1} \rangle \\ & = \phi \langle \nabla h(\bar{z}_k) - \nabla h(z_k), z_k - z_{k+1} \rangle \\ & = \phi [D_h(z_{k+1}, \bar{z}_k) - D_h(z_{k+1}, z_k) - D_h(z_k, \bar{z}_k)]. \end{aligned} \quad (39)$$

Now, multiplying both sides of (39) by $\frac{\lambda_k}{\lambda_{k-1}}$ then gives

$$\begin{aligned} & \lambda_k(\langle F(z_{k-1}), z_k - z_{k+1} \rangle + g(z_k) - g(z_{k+1})) \\ & \leq \theta_k [D_h(z_{k+1}, \bar{z}_k) - D_h(z_{k+1}, z_k) - D_h(z_k, \bar{z}_k)]. \end{aligned} \quad (40)$$

Let $z^* \in \Omega$. Setting $z = z^*$ in (38), summing with (39) and rearranging yields

$$\begin{aligned} & \lambda_k(\langle F(z_k), z_k - z^* \rangle + g(z_k) - g(z^*)) \\ & \leq D_h(z^*, \bar{z}_k) - D_h(z^*, z_{k+1}) - D_h(z_{k+1}, \bar{z}_k) \\ & \quad + \theta_k [D_h(z_{k+1}, \bar{z}_k) - D_h(z_{k+1}, z_k) - D_h(z_k, \bar{z}_k)] \\ & \quad + \lambda_k(F(z_k) - F(z_{k-1}), z_k - z_{k+1}). \end{aligned} \quad (41)$$

We observe that the left side of (41) is nonnegative as a consequence of (10) and A.3:

$$0 \leq \langle F(z^*), z_k - z^* \rangle + g(z_k) - g(z^*) \leq \langle F(z_k), z_k - z^* \rangle + g(z_k) - g(z^*). \quad (42)$$

To estimate the final term of (41), we use the Cauchy–Schwarz inequality, the local Lipschitz estimate (36) and σ -strong convexity of h to obtain

$$\begin{aligned} & \lambda_k \langle F(z_k) - F(z_{k-1}), z_k - z_{k+1} \rangle \\ & \leq \frac{\sqrt{\sigma \theta_k \theta_{k-1}}}{2} \|z_k - z_{k-1}\| \|z_k - z_{k+1}\| \\ & \leq \frac{\sigma \theta_{k-1}}{4} \|z_k - z_{k-1}\|^2 + \frac{\sigma \theta_k}{4} \|z_k - z_{k+1}\|^2 \\ & \leq \frac{\theta_{k-1}}{2} D_h(z_k, z_{k-1}) + \frac{\theta_k}{2} D_h(z_{k+1}, z_k). \end{aligned} \quad (43)$$

Combining (41), (42) and (43) gives

$$\begin{aligned} D_h(z^*, z_{k+1}) & \leq D_h(z^*, \bar{z}_k) + (\theta_k - 1) D_h(z_{k+1}, \bar{z}_k) - \theta_k D_h(z_k, \bar{z}_k) \\ & \quad - \frac{\theta_k}{2} D_h(z_{k+1}, z_k) + \frac{\theta_{k-1}}{2} D_h(z_k, z_{k-1}). \end{aligned} \quad (44)$$

Now, applying Proposition 2.1(c) with $\nabla h(z_{k+1}) = \frac{\phi \nabla h(\bar{z}_{k+1}) - \nabla h(\bar{z}_k)}{\phi - 1}$ and rearranging yield

$$\begin{aligned} \frac{\phi}{\phi - 1} D_h(z^*, \bar{z}_{k+1}) & = D_h(z^*, z_{k+1}) + \frac{\phi}{\phi - 1} D_h(z_{k+1}, \bar{z}_{k+1}) \\ & \quad + \frac{1}{\phi - 1} \left[D_h(z^*, \bar{z}_k) - D_h(z_{k+1}, \bar{z}_k) \right]. \end{aligned} \quad (45)$$

Combining (44) and (45), followed by collecting like-terms and rearranging, gives

$$\begin{aligned} & \frac{\phi}{\phi - 1} D_h(z^*, \bar{z}_{k+1}) + \frac{\theta_k}{2} D_h(z_{k+1}, z_k) - D_h(z_{k+1}, \bar{z}_{k+1}) \\ & \leq \frac{\phi}{\phi - 1} D_h(z^*, \bar{z}_k) + \frac{\theta_{k-1}}{2} D_h(z_k, z_{k-1}) - \theta_k D_h(z_k, \bar{z}_k) \\ & \quad + \left(\theta_k - \frac{\phi}{\phi - 1} \right) D_h(z_{k+1}, \bar{z}_k) + \frac{1}{\phi - 1} D_h(z_{k+1}, \bar{z}_{k+1}). \end{aligned} \quad (46)$$

Next we apply Proposition 2.1(c) once again to see that

$$\begin{aligned} D_h(z_{k+1}, \bar{z}_{k+1}) &= \frac{\phi - 1}{\phi} [D_h(z_{k+1}, z_{k+1}) - D_h(\bar{z}_{k+1}, z_{k+1})] \\ &\quad + \frac{1}{\phi} [D_h(z_{k+1}, \bar{z}_k) - D_h(\bar{z}_{k+1}, \bar{z}_k)] \\ &\leq \frac{1}{\phi} D_h(z_{k+1}, \bar{z}_k). \end{aligned} \quad (47)$$

The final line of (46) can therefore be estimated as

$$\begin{aligned} \left(\theta_k - \frac{\phi}{\phi - 1} \right) D_h(z_{k+1}, \bar{z}_k) + \frac{1}{\phi - 1} D_h(z_{k+1}, \bar{z}_{k+1}) \\ \leq \left(\theta_k - \frac{\phi}{\phi - 1} + \frac{1/\phi}{\phi - 1} \right) D_h(z_{k+1}, \bar{z}_k) \\ = \left(\theta_k - 1 - \frac{1}{\phi} \right) D_h(z_{k+1}, \bar{z}_k). \end{aligned} \quad (48)$$

Substituting (48) into (46) gives (37), which completes the proof. $\square$

The following is our main result regarding convergence of the Bregman-adaptive GRAAL.

Theorem 4.1 *Suppose Assumptions A.1–A.4 hold, and F is locally Lipschitz continuous on the bounded set $\text{dom } g \cap \text{int dom } h$. Choose $\phi \in (1, \varphi)$ and $\rho \in \left[1, \frac{1}{\phi} + \frac{1}{\phi^2}\right)$. Then, the sequences (z_k) and $(\bar{z}_k)$ generated by Algorithm 3 converge to a point in Ω whenever $\lambda_{\max} > 0$ is sufficiently small.*

Proof Since $\text{dom } g \cap \text{int dom } h$ is bounded and F is locally Lipschitz, there exists $L > 0$ such that F is L -Lipschitz continuous on $\text{dom } g \cap \text{int dom } h$. Suppose $\lambda_{\max}$ is sufficiently small so that $\lambda_{\max} \leq \frac{\phi\sqrt{\phi\sigma}}{2L}$ holds. Then, applying Lemma 4.2 gives $\theta_k \geq \frac{\sigma\phi^3}{4L^2\lambda_{\max}^2} \geq 1$ for all $k \in \mathbb{N}$. Also, since $\theta_k \leq \rho\phi < 1 + \frac{1}{\phi}$, there exists an $\epsilon > 0$ such that $\theta_k - 1 - \frac{1}{\phi} \leq -\epsilon$ for all $k \in \mathbb{N}$.

Now, let $z^* \in \Omega$ be arbitrary and denote by (η_k) the sequence

$$\eta_k := \frac{\phi}{\phi - 1} D_h(z^*, \bar{z}_k) + \frac{\theta_{k-1}}{2} D_h(z_k, z_{k-1}) - D_h(z_k, \bar{z}_k).$$

Applying Lemma 4.3 yields

$$\eta_{k+1} \leq \eta_k - \epsilon D_h(z_{k+1}, \bar{z}_k). \quad (49)$$

Next, we should show that (η_k) is bounded below. To this end, using the three-point identity (Proposition 2.1(a)) followed by (38) gives

$$\begin{aligned}
 & D_h(z^*, z_{k+1}) + D_h(z_{k+1}, \bar{z}_{k+1}) \\
 &= D_h(z^*, \bar{z}_{k+1}) + \langle \nabla h(\bar{z}_{k+1}) - \nabla h(z_{k+1}), z^* - z_{k+1} \rangle \\
 &= D_h(z^*, \bar{z}_{k+1}) + \frac{1}{\phi} \langle \nabla h(\bar{z}_k) - \nabla h(z_{k+1}), z^* - z_{k+1} \rangle \\
 &\leq D_h(z^*, \bar{z}_{k+1}) + \frac{\lambda_k}{\phi} (\langle F(z_k), z^* - z_{k+1} \rangle - g(z_{k+1}) + g(z^*)) \quad (50) \\
 &= D_h(z^*, \bar{z}_{k+1}) + \frac{\lambda_k}{\phi} \langle F(z_k) - F(z_{k+1}), z^* - z_{k+1} \rangle \\
 &\quad + \frac{\lambda_k}{\phi} (\langle F(z_{k+1}), z^* - z_{k+1} \rangle - g(z_{k+1}) + g(z^*)).
 \end{aligned}$$

Now, as with (16), the final term in (50) is nonnegative. Since (λ_k) and (θ_k) are bounded and separated from 0 by Lemma 4.2, there exists a constant $M > 0$ such that

$$\begin{aligned}
 & \lambda_k \langle F(z_k) - F(z_{k+1}), z^* - z_{k+1} \rangle \\
 &\leq \frac{\lambda_k}{\lambda_{k+1}} \lambda_{k+1} \|F(z_k) - F(z_{k+1})\| \|z^* - z_{k+1}\| \quad (51) \\
 &\leq \frac{\lambda_k}{\lambda_{k+1}} \frac{\sigma \sqrt{\theta_{k+1} \theta_k}}{2} \|z_k - z_{k+1}\| \|z^* - z_{k+1}\| \leq M.
 \end{aligned}$$

Combining (50) and (51), noting that $\phi < \frac{\phi}{\phi-1}$ then gives

$$\begin{aligned}
 D_h(z_{k+1}, \bar{z}_{k+1}) &\leq D_h(z^*, \bar{z}_{k+1}) + \frac{M}{\phi} \\
 &\leq \phi D_h(z^*, \bar{z}_{k+1}) + M \implies \eta_{k+1} \geq -M, \quad (52)
 \end{aligned}$$

which establishes that (η_k) is bounded below.

Next, by telescoping the inequality (49), deduce that (η_k) is bounded and $D_h(z_{k+1}, \bar{z}_k) \rightarrow 0$ as $k \rightarrow \infty$. Referring to (47), it follows that $D_h(z_{k+1}, \bar{z}_{k+1}) \rightarrow 0$. Also, by applying Proposition 2.1(c) with the identity $\nabla h(z_k) = \frac{\phi \nabla h(\bar{z}_k) - \nabla h(\bar{z}_{k-1})}{\phi-1}$, we obtain

$$\begin{aligned}
 D_h(z_{k+1}, z_k) &= \frac{\phi}{\phi-1} [D_h(z_{k+1}, \bar{z}_k) - D_h(z_k, \bar{z}_k)] \\
 &\quad - \frac{1}{\phi-1} [D_h(z_{k+1}, \bar{z}_{k-1}) - D_h(z_k, \bar{z}_{k-1})] \\
 &\leq \frac{\phi}{\phi-1} D_h(z_{k+1}, \bar{z}_k) + \frac{1}{\phi-1} D_h(z_k, \bar{z}_{k-1}) \rightarrow 0.
 \end{aligned}$$

Altogether, noting that (θ_k) is bounded, we deduce that

$$\lim_{k \rightarrow \infty} \eta_k = \frac{\phi}{\phi - 1} \lim_{k \rightarrow \infty} D_h(z^*, \bar{z}_k),$$

and so, in particular, $\lim_{k \rightarrow \infty} D_h(z^*, \bar{z}_k)$ exists.

Next, using σ -strong convexity of h , we deduce that $z_{k+1} - \bar{z}_k \rightarrow 0$. Let $\bar{z} \in \mathcal{H}$ be a cluster point of the bounded sequence $(\bar{z}_k)$. Then, there exists a subsequence $(\bar{z}_{k_j})$ such that $\bar{z}_{k_j} \rightarrow \bar{z}$ and $z_{k_j+1} \rightarrow \bar{z}$ as $j \rightarrow \infty$. Now, recalling (38) gives

$$\begin{aligned} & \lambda_k (\langle F(z_{k_j}), z_{k_j+1} - z \rangle + g(z_{k_j+1}) - g(z)) \\ & \leq \langle \nabla h(\bar{z}_{k_j}) - \nabla h(z_{k_j+1}), z_{k_j+1} - z \rangle \quad \forall z \in \mathcal{H}, \end{aligned}$$

and taking the limit-infimum of both sides as $j \rightarrow \infty$ shows that $\bar{z} \in \Omega$. Since $z^* \in \Omega$ was chosen in Lemma 4.3 to be arbitrary, we can now set $z^* = \bar{z}$. It then follows that $\lim_{j \rightarrow \infty} D_h(z^*, \bar{z}_{k_j}) = 0$, and consequently, $\lim_{j \rightarrow \infty} \eta_{k_j} = 0$. Also note that for $n \geq k_j$, we have $\eta_n \leq \eta_{k_j}$ from Lemma 3.2, and therefore,

$$\frac{\phi}{\phi - 1} \lim_{n \rightarrow \infty} D_h(z^*, \bar{z}_n) = \lim_{n \rightarrow \infty} \eta_n \leq \lim_{j \rightarrow \infty} \eta_{k_j} = 0,$$

and $\bar{z}_k \rightarrow z^*$ from strong convexity. The fact that $z_k \rightarrow z^*$ follows since $z_k - \bar{z}_k \rightarrow 0$. $\square$

Remark 4.1 The energy (49) used in the proof of Theorem 4.1 is not completely analogous to the one used in the fixed step-size case given in Lemma 3.2. Notably, the coefficient of the final term of η_k in (49) has coefficient -1 , whereas the corresponding term in Lemma 3.2 has coefficient $-\phi$. Although it is unlikely to be useful in practice, another interesting feature of the proof is that it requires the maximum step size to satisfy the upper bound

$$\lambda_{\max} \leq \frac{\phi \sqrt{\phi \sigma}}{2L} = \sqrt{\frac{\phi}{\sigma}} \frac{\sigma \phi}{2L}.$$

Note that when $\phi > \sigma$, this upper bound is looser than the upper bound of $\lambda < \frac{\sigma \phi}{2L}$ required for B-GRAAL (Algorithm 2). Thus, it is possible for maximum step size of B-aGRAAL to be larger than the step size required for B-GRAAL. This is the case, for instance, if D_h is the KL divergence on the simplex in which case $\sigma = 1 < \phi$.

Also, the proof of Theorem 4.1 required L be a Lipschitz constant for F . However, thanks to Lemma 4.2, it would have sufficed for L to satisfy the weaker Lipschitz-like inequality $\|F(z_k) - F(z_{k-1})\| \leq L \|z_k - z_{k-1}\|$ for all $k \in \mathbb{N}$. In turn, this would allow for a greater upper bound for $\lambda_{\max}$ which is inversely related to L .

Remark 4.2 A similar analysis could be applied to infer weak convergence in infinite-dimensional spaces for Algorithms 2 and 3. However, this would firstly require the additional condition of weak-weak continuity in ∇h . Lemma 4.2 in this context would

also require that F is Lipschitz continuous over bounded sets, which in infinite dimensions is a stronger condition than local Lipschitz continuity. We therefore decided to work in finite dimensions for simplicity.

5 Numerical Experiments

In this section, we present some experimental comparisons between Algorithms 2 and 3 and their respective Euclidean counterparts. To be precise, in Sect. 5.1 we compare the *Kullback–Leibler* version of Algorithm 2 to the Euclidean version, which is equivalent to GRAAL. In Sect. 5.2, we make the same comparison for Algorithm 3, recalling once again that aGRAAL is recovered in the Euclidean case. Finally, in Sect. 5.3, we compare the Euclidean aGRAAL to Algorithm 3 with two distinct choices of Bregman distances. All experiments are run in Python 3 on a Windows 10 machine with 8GB memory and an Intel(R) Core(TM) i7-10510U CPU @ 1.80GHz processor.

We consider three different problems, all of which can be formulated as the variational inequality (10) for different choices of the function g and the operator F . Note that the solutions of the variational inequality (10) can also be characterised as the monotone inclusion

$$\text{find } z^* \in \mathcal{H} \text{ such that } 0 \in F(z^*) + \partial g(z^*).$$

Thus, by noting that (35) implies $0 \in F(z_k) + \partial g(z_{k+1}) + \frac{1}{\lambda_k} (\nabla h(z_{k+1}) - \nabla h(\bar{z}_k))$, we monitor the quantity J_k given by

$$J_k := \frac{1}{\lambda_k} (\nabla h(\bar{z}_k) - \nabla h(z_{k+1})) + F(z_{k+1}) - F(z_k) \in F(z_{k+1}) + \partial g(z_{k+1}) \quad (53)$$

as a natural residual for Algorithm 3. The analogous expression for Algorithm 2 is given by replacing λ_k in (53) with λ .

In all experiments, we run each algorithm on for the same (fixed) number of iterations on 10 random instances of the chosen problem. The figures in each section show the decrease in residual over time, with each faint line representing one instance and the bold line showing the mean behaviour. We use the parameters $\phi = 1.5$, $\lambda_{\max} = 10^6$ throughout, and initial step size and iterates as described in each section.

5.1 Matrix Games

To test the fixed step-size B-GRAAL (Algorithm 2), we first consider the following matrix game between two players

$$\min_{x \in \Delta^n} \max_{y \in \Delta^n} \langle Mx, y \rangle, \quad (54)$$

where $\Delta^n := \{x \in \mathbb{R}_+^n : \sum_{i=1}^n x_i = 1\}$ denotes the unit simplex and $M \in \mathbb{R}^{n \times n}$ a given matrix. This problem is of the form specified in (1) and so can be formulated as the variational inequality (10).

In particular, we consider the specific problem in the form of (54) of placing a server on a network $G = (V, E)$ with n vertices in a way that minimises its response time. In this problem, a request will originate at some vertex $v_j \in V$, which is not known ahead of time, and the objective is to place the server at a vertex $v_i \in V$ such the response time, as measured by the graphical distance $d(v_i, v_j)$, is minimised. We consider the case where the request location $v_j \in V$ is a decision made by an adversary. The decision variable $x \in \Delta^n$ (resp. $y \in \Delta^n$) models mixed strategies for placement of the server (resp. request origin). In other words, x_t (resp. y_t) is the probability of the server (resp. request origin) being located at the node v_t for $t = 1, \dots, n$. The matrix M is taken to be the distance matrix of the graph G , that is, $M_{ij} = d(v_i, v_j)$ for all vertices $v_i, v_j \in V$. In this way, the objective function in (54) measures the expected response time, which we would like to minimise while our adversary seeks to maximise it.

We compare Algorithm 2 with the squared norm $h(z) = \frac{1}{2}\|z\|^2$, which generates the squared Euclidean distance $D_h(u, v) = \frac{1}{2}\|u - v\|^2$, and the negative entropy $h(z) = \sum_{i=1}^{2n} z_i \log z_i$, which generates the KL divergence $D_h(u, v) = \sum_{i=1}^{2n} u_i \log \frac{u_i}{v_i} + v_i - u_i$, where $z = (x, y) \in \mathbb{R}^{2n}$ is the concatenation of the two vectors x and y . Both of these choices for h are 1-strongly convex on Δ^n . A potential advantage of the KL projection onto the simplex is that it has the simple closed-form expression $x \mapsto \frac{x}{\|x\|_1}$, whereas the Euclidean projection has no closed form and takes $O(n \log n)$ time (see, for instance, [18, 19, 50]). We run two experiments with $n = 500$ and $n = 1000$, and the results are shown, respectively, in Figs. 1 and 2. Initial points are chosen as $x_0 = y_0 = (\frac{1}{n}, \dots, \frac{1}{n}) \in \Delta^n$, and $z_0 = (x_0, y_0)$, then $\bar{z}_0$ as a random perturbation of z_0 . The Lipschitz constant of F is computed as $L = \|M\|_2$, and the algorithm step size is taken as $\lambda = \frac{\varphi}{2L}$. Under these conditions, convergence to a solution is guaranteed by Theorem 3.1.

The residual, given by $\min_{i \leq k} \|J_i\|^2$, is shown in Figs. 1a and 2a. The time per iteration is also shown in Figs. 1b and 2b. Despite the KL projection being faster than the Euclidean projection, overall, the Euclidean method performed better in this instance.

We now move onto experiments for the adaptive Algorithm 3.

5.2 Gaussian Communication

We now turn our attention to maximising the information capacity of a noisy *Gaussian communication channel* [16, Chapter 9]. In this problem, the goal is to allocate a total power of P across m channels, represented by $p \in \mathbb{R}_+^m$, to maximise the total information capacity of the channels in the presence of allocated noise, represented by $n \in \mathbb{R}_+^m$. The information capacity of the i th channel, denoted $C_i(p_i, n_i)$, is the

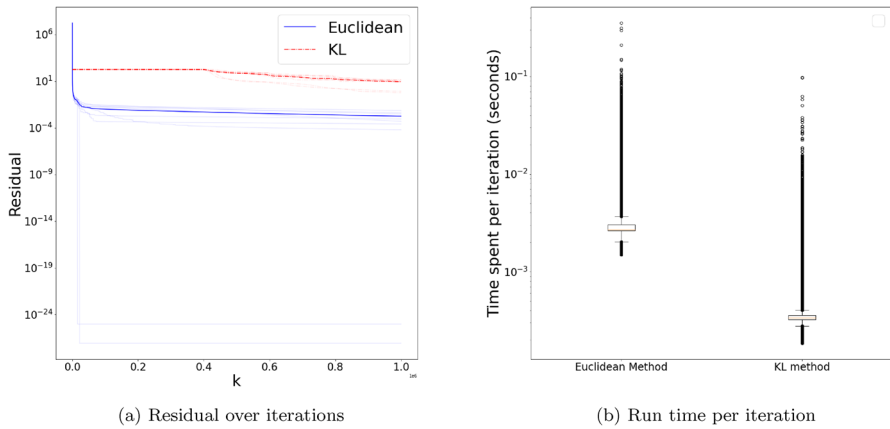


Fig. 1 Matrix game results for $n = 500$

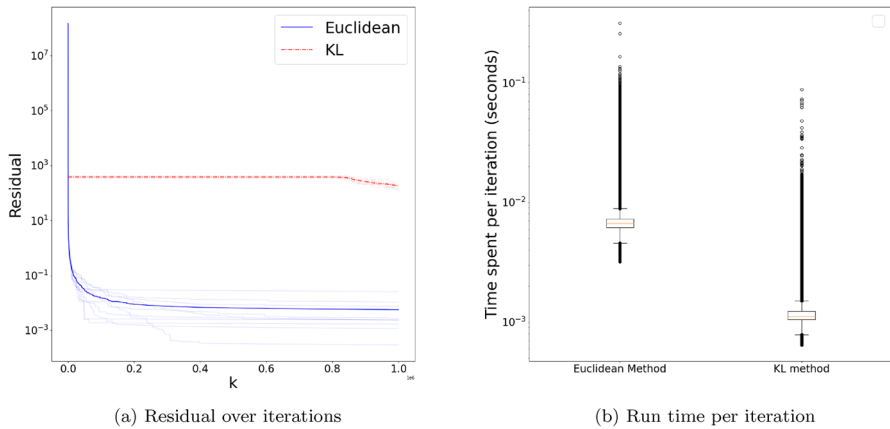


Fig. 2 Matrix game results for $n = 1000$

function of power p_i and noise level n_i and is given by

$$C_i(p_i, n_i) = \log \left(1 + \frac{\beta_i p_i}{\mu_i + n_i} \right),$$

where $\mu_i > 0$ and $\beta_i > 0$ are given constants.

Assuming a total power level of P and a total noise level of N , optimising for the worst-case scenario by treating the noise allocation as an adversary gives the convex-concave game

$$\max_{p \in \Delta_P^m} \min_{n \in \Delta_N^m} \sum_{i=1}^m C_i(p_i, n_i),$$

where $\Delta_T^n := \{x \in \mathbb{R}_+^n : \sum_{i=1}^n x_i = T\}$ is a *scaled* simplex. This problem is also of the form specified in (1) and so can also be formulated as the variational inequality (10).

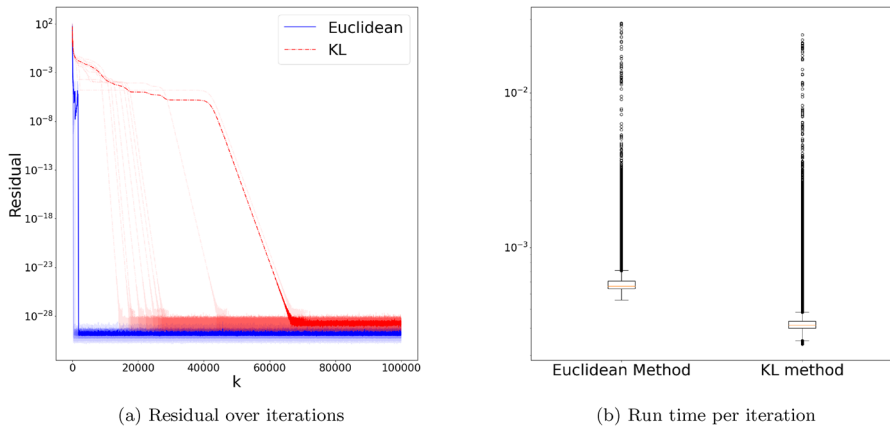


Fig. 3 Gaussian communication channel results for $m = 100$

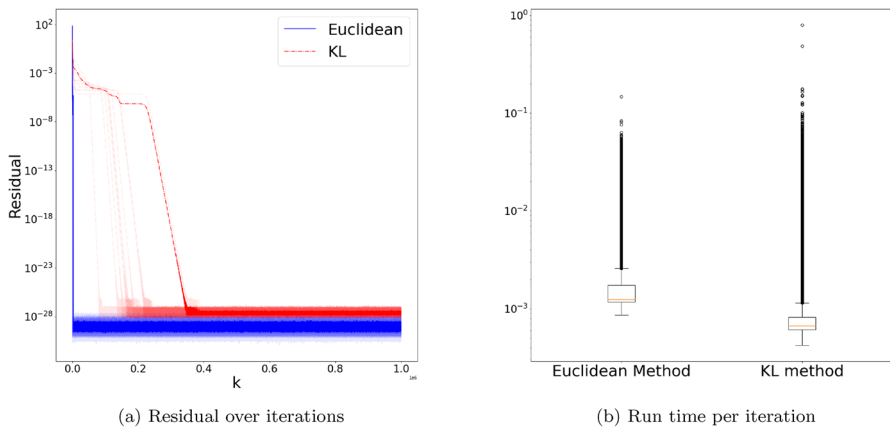


Fig. 4 Gaussian communication channel results for $m = 200$

The Lipschitz constant of the operator F for this problem is not straightforward to compute, so we apply the adaptive algorithms. Similarly to Sect. 5.1, we compare the Euclidean and KL versions of Algorithm 3. Since $x \mapsto x \log x$ is $\frac{1}{M}$ -strongly convex for $0 < x \leq M$, the strong convexity constant of the negative entropy over $\Delta_P^m \times \Delta_N^m$ is $\min\{\frac{1}{P}, \frac{1}{N}\}$. In our experiments, we set $(P, N) = (500, 50)$ and generate $\beta \in (0, P]^m$ and $\mu \in (1, N + 1]^m$ uniformly. The initial points are chosen as $p_0 = (\frac{P}{m}, \dots, \frac{P}{m}) \in \Delta_P^m$, $n_0 = (\frac{N}{m}, \dots, \frac{N}{m}) \in \Delta_N^m$, and $z_0 = (p_0, n_0)$, and the initial step size is taken as $\lambda_0 = \frac{\|z_0 - \bar{z}_0\|^2}{\|F(z_0) - F(\bar{z}_0)\|^2}$ where $\bar{z}_0$ is again a small random perturbation of z_0 . It is also worth noting that for the KL version, we had to multiply λ_0 by a small constant (10^{-2}) to avoid numerical instability issues.

We run two experiments, for $m = 100$ and $m = 200$, and plot the results in Figs. 3 and 4, respectively. The KL method is slower than the Euclidean method in terms of the number of iterations and time. However, unlike in the previous section, both methods reach a similar final accuracy.

5.3 Cournot Completion

Our final example is a standard N -player Cournot oligopoly model [14, Example 2.1]. This is a system in which N independent firms supply the market with a quantity of some common good or service. More formally, each firm seeks to maximise their utility subject to their capacity, that is,

$$\max_{0 \leq x_i \leq C_i} u_i(x_i, x_{-i}) = x_i P(x_T) - c_i x_i, \quad (55)$$

where

- $x_i \geq 0$ is the quantity of the good supplied by the i th firm, $i = 1, \dots, N$.
- $x_{-i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_N)$ is the quantity of the good supplied by all other firms.
- $x_T := \sum_{i=1}^N x_i$ is the total amount supplied.
- $C_i > 0$ is the production capacity of the i th firm.
- $c_i > 0$ is the production cost of the i th firm.
- $P: \mathbb{R}_+ \rightarrow \mathbb{R}$ is the inverse demand curve.

In this section, we consider solutions to this problem in the sense of *Nash equilibria*, which is equivalent to the variational inequality (10) with

$$F = \left(-\frac{\partial u_1}{\partial x_1}, \dots, -\frac{\partial u_N}{\partial x_N} \right), \quad K = [0, C_1] \times \dots \times [0, C_N], \quad g = \iota_K.$$

By choosing a function h such that $\text{dom } h = K$, it is possible to implicitly enforce the capacity constraint in this problem and avoid performing projections. Two examples, over a single closed interval $[\alpha, \beta]$, present themselves which satisfy our assumptions:

$$h_1(x) = (x - \alpha) \log(x - \alpha) + (\beta - x) \log(\beta - x), \quad h_2(x) = -\sqrt{(x - \alpha)(\beta - x)}.$$

When $\alpha = 0$ and $\beta = 1$, h_1 is called the *Fermi-Dirac Entropy*, and similarly when $\alpha = -1$ and $\beta = 1$, D_{h_2} is called the *Hellinger distance* ([9, Example 2.2], [8, Example 1], [49, Example 2.1]). Then, we sum over the independent functions of one variable, with appropriately set intervals $\alpha = 0$, $\beta = C_i$, to create the Bregman distance over the closed box K . We also compute the strong convexity constants by minimising $\nabla^2 h_1$ over (α, β) , which gives $\sigma = \frac{4}{\beta - \alpha}$, and similarly for h_2 , $\sigma = \frac{2}{\beta - \alpha}$.

In our experiments, P is taken to be a linear inverse demand curve given by $P(x) = a - bx$ for $a, b > 0$. All parameters are generated by a log-normal distribution, except the cost vector $c \in \mathbb{R}^N$, which is generated uniformly in $\left[\frac{C_1}{100}, \frac{C_1}{5} \right] \times \dots \times \left[\frac{C_N}{100}, \frac{C_N}{5} \right]$. We run two experiments, with $N = 2000$ and $N = 5000$, the results of which are shown in Figs. 5 and 6. The initial points are chosen as $z_0 = \frac{C}{2}$, and $\lambda_0 = 1$.

We observe here that the final accuracy depends heavily on the choice of Bregman distance. In both figures, we observe that the Hellinger method is not making any progress after approximately 200 iterations, while the Euclidean method achieves a modest final accuracy. Meanwhile, the Bit method is very fast and accurate, converging

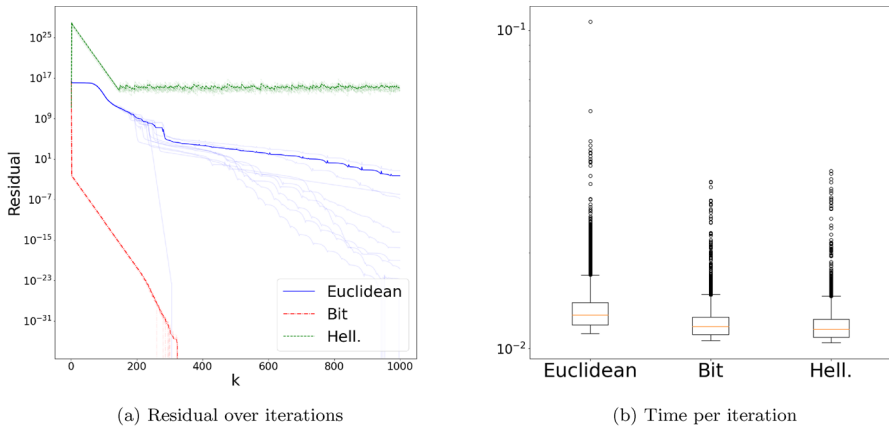


Fig. 5 Results for $N = 2000$

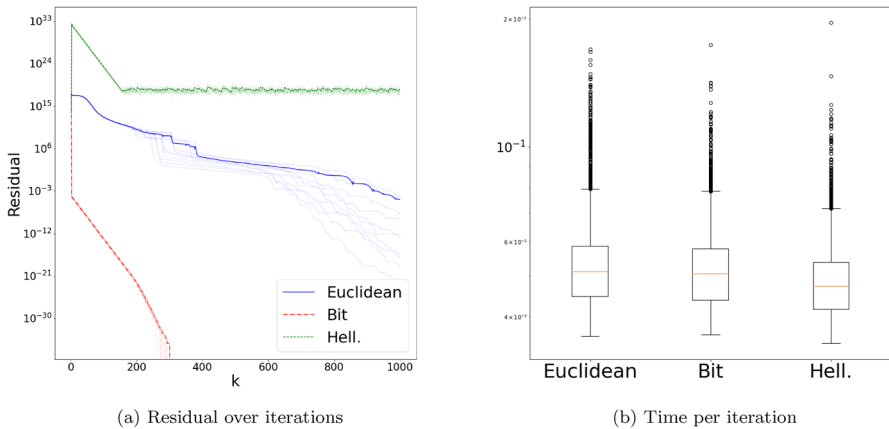


Fig. 6 Results for $N = 5000$

to a near 0 tolerance in roughly 300 iterations in all instances. Finally, we note that all methods are roughly equal in terms of time per iteration. This is to be expected, since the Euclidean method requires a projection which evaluates $\min\{\max\{x_i, 0\}, C_i\}$ for each component, whereas the Bregman methods require evaluating $\nabla h, (\nabla h)^{-1}$ over each component—all of which are taking $O(N)$ time.

6 Conclusion

In this paper, we extended the adaptive method aGRAAL (Algorithm 1) to the Bregman distance setting. We proposed two such extensions: The first, Algorithm 2, generalises the fixed step-size GRAAL and converges under the same assumptions for a strongly convex Bregman function $h: \mathcal{H} \rightarrow \mathbb{R}$. The second, Algorithm 3, generalises Algorithm 1 and converges in a more restrictive setting. We first examined the per-

formance of Algorithm 2 and found that the KL version is less favourable than the Euclidean version, despite the reduced time per iteration. We then tested Algorithm 3 on a convex–concave game for Gaussian communication channels, where our new method performed worse with respect to the KL divergence when compared to the Euclidean method, although the run-time per iteration was again significantly shorter as was expected. Finally, we examined a Cournot completion model, where one of Bregman based methods reached a much higher accuracy very quickly.

We conclude by outlining directions for further research:

- It would be interesting to know whether Algorithm 3 can be shown to converge in a more general setting than what we have shown, and if so, under what circumstances. The difficulties in our analysis arose from two issues: first, the estimate derived in (47) for the Bregman case is weaker than the Euclidean equality $\|z_{k+1} - \bar{z}_{k+1}\|^2 = \frac{1}{\phi^2} \|z_{k+1} - \bar{z}_k\|^2$ used in [40], and second, the inability to bound θ_k below without additional assumptions (as the bound in Lemma 4.2 can be arbitrarily small in general).
- Throughout this paper, h was assumed strongly convex, but whether or not this assumption can be relaxed is unclear. One potential replacement for strong convexity is considered in [6, 36] where h is twice differentiable such that the Hessian matrix $\nabla^2 h(z)$ is positive definite for all $z \in \text{int dom } h$. Within the scope of twice differentiable functions, such a condition lies between strict and strong convexity, the main consequence being that ∇h is locally Lipschitz and h is locally strongly convex. Indeed, for σ -strongly convex h with η -Lipschitz gradient, one can derive the estimates

$$\frac{\sigma}{2} \|z - z'\|^2 \leq D_h(z, z') \leq \frac{\eta}{2} \|z - z'\|^2 \quad \forall z, z'.$$

If such inequalities hold on a local scale, then it would remain to be seen whether these coefficients can be estimated and used in the same way that λ_k approximates an inverse of the local Lipschitz constant of F .

- In the context of the convex composite optimisation problems of the form $\min_{x \in \mathcal{H}} f(x) + g(x)$ where g is non-smooth and f is smooth, the Bregman proximal gradient algorithm [8, 49] is known to converge in a more general setting than Lipschitz continuity. Specifically, L -Lipschitz continuity of ∇f can be relaxed to convexity of the function $Lh - f$, which indeed holds if ∇f is Lipschitz and h is strongly convex. It would be interesting to see if a similar relaxation of Lipschitz continuous can be used in the context of the algorithms discussed here.

Acknowledgements DJU and MKT are supported in part by Australian Research Council Grant DE200100063. The authors thank the anonymous referees for their valuable comments which helped to improve the manuscript.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included

in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Adolphs, L., Daneshmand, H., Lucchi, A., Hofmann, T.: Local saddle point optimization: A curvature exploitation approach. In: The 22nd International Conference on Artificial Intelligence and Statistics, pp. 486–495. PMLR (2019)
2. Akimoto, Y.: Saddle point optimization with approximate minimization oracle. In: Proceedings of the Genetic and Evolutionary Computation Conference, pp. 493–501. ACM (2021)
3. Alacaoglu, A., Malitsky, Y., Cevher, V.: Convergence of adaptive algorithms for weakly convex constrained optimization. In: Advances in Neural Information Processing Systems (2021)
4. Bauschke, H.H., Borwein, J.M.: Legendre functions and the method of random Bregman projections. *J. Convex Anal.* **4**(1), 27–67 (1997)
5. Bauschke, H.H., Combettes, P.L.: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, vol. 408. Springer, New York (2011)
6. Bauschke, H.H., Lewis, A.S.: Dykstra's algorithm with Bregman projections: a convergence proof. *Optimization* **48**, 409–427 (2000)
7. Bauschke, H.H., Borwein, J.M., Combettes, P.L.: Bregman monotone optimization algorithms. *SIAM J. Control. Optim.* **42**(2), 596–636 (2003)
8. Bauschke, H.H., Bolte, J., Teboulle, M.: A descent Lemma beyond Lipschitz gradient continuity: first-order methods revisited and applications. *Math. Oper. Res.* **42**(2), 330–348 (2017)
9. Bauschke, H.H., Bolte, J., Chen, J., Teboulle, M., Wang, X.: On linear convergence of non-Euclidean gradient methods without strong convexity and Lipschitz gradient continuity. *J. Optim. Theory Appl.* **182**, 09 (2019)
10. Beck, A., Teboulle, M.: Mirror descent and nonlinear projected subgradient methods for convex optimization. *Oper. Res. Lett.* **31**(3), 167–175 (2003)
11. Bello Cruz, J.Y., Díaz Millán, R.: A variant of forward-backward splitting method for the sum of two monotone operators with a new search strategy. *Optimization* **64**(7), 1471–1486 (2015)
12. Borwein, J.M., Vanderwerff, J.D.: *Convex Functions: Constructions, Characterizations and Counterexamples*, vol. 172. Cambridge University Press, Cambridge (2010)
13. Borwein, J.M., Reich, S., Sabach, S.: A characterization of Bregman firmly nonexpansive operators using a new monotonicity concept. *J. Nonlinear Convex Anal.* **12**(1), 161–184 (2011)
14. Bravo, M., Leslie, D., Mertikopoulos: Bandit learning in concave N -person games. In: 32nd Conference on Neural Information Processing Systems, Montréal (2018)
15. Censor, Y., Iusem, A., Zenios, S.: An interior point method with Bregman functions for the variational inequality problem with paramonotone operators. *Math. Program.* **81**, 373–400 (1998)
16. Cover, T.M., Thomas, J.A.: *Elements of information theory*. Wiley-Interscience, New York (2006)
17. Chen, G., Teboulle, M.: Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM J. Optim.* **3**(3), 538–543 (1993)
18. Chen, Y., Ye, X.: Projection onto a simplex (2011). [arXiv:1101.6081](https://arxiv.org/abs/1101.6081)
19. Dai, Y., Chen, C.: Distributed projections onto a simplex (2022). [arXiv:2204.08153](https://arxiv.org/abs/2204.08153)
20. Dürr, H.B., Zeng, C., Ebenbauer, C.: Saddle point seeking for convex optimization problems. In: 9th IFAC Symposium on Nonlinear Control Systems, vol. 46, pp. 540–545 (2013)
21. Fukushima, M.: Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality. *Math. Program.* **53**, 01 (1992)
22. Gibali, A.: A new Bregman projection method for solving variational inequalities in Hilbert spaces. *Pure Appl. Funct. Anal.* **3**(3), 403–415 (2018)
23. Gibali, A., Jolaoso, L.O., Mewomo, O.T., Taiwo, A.: Fast and simple Bregman projection methods for solving variational inequalities and related problems in Banach spaces. *Results Math.* **75**(4), 1–36 (2020)
24. Guo, K., Zhu, C.: On the linear convergence of a Bregman proximal point algorithm. *J. Nonlinear Var. Anal.* **6**(2), 5–14 (2022)

25. Hamedani, E.Y., Aybat, N.S.: A primal-dual algorithm with line search for general convex-concave saddle point problems. *SIAM J. Optim.* **31**(2), 1299–1329 (2021)
26. Hieu, D.V., Chalamjiak, P.: Modified extragradient method with Bregman distance for variational inequalities. *Appl. Anal.* **101**(2), 655–670 (2022)
27. Hieu, D.V., Reich, S.: Two Bregman projection methods for solving variational inequalities. *Optimization* **71**(7), 1777–1802 (2022)
28. Izuchukwu, C., Shehu, Y., Yao, J.C.: New inertial forward-backward type for variational inequalities with Quasi-monotonicity. *J. Glob. Optim.* **84**, 441–464 (2022)
29. Iusem, A.N., Svaiter, B.F.: A variant of Korpelevich's method for variational inequalities with a new search strategy. *Optimization* **42**(4), 309–321 (1997)
30. Jolaoso, L.O., Aphane, M.: Weak and strong convergence Bregman extragradient schemes for solving pseudo-monotone and non-Lipschitz variational inequalities. *J. Inequal. Appl.* **2020**(1), 1–25 (2020)
31. Jolaoso, L.O., Shehu, Y.: Single Bregman projection method for solving variational inequalities in reflexive Banach spaces. *Appl. Anal.* 1–22 (2021)
32. Jolaoso, L.O., Aphane, M., Khan, S.H.: Two Bregman projection methods for solving variational inequality problems in Hilbert spaces with applications to signal processing. *Symmetry* **12**(12), 2007 (2020)
33. Kakade, S., Shalev-Shwartz, S., Tewari, A.: On the duality of strong convexity and strong smoothness: Learning applications and matrix regularization. Technical report, Toyota Technological Institute (2009)
34. Korpelevič, G.M.: The extragradient method for finding saddle points and other problems. *Ekonomika i Matematicheskie Metody* **12**, 747–756 (1976)
35. Krichene, W., Bayen, A., Bartlett, P.L.: Accelerated mirror descent in continuous and discrete time. In: *Advances in Neural Information Processing Systems*, vol. 28. Curran Associates, Inc. (2015)
36. Laude, E., Ochs, P., Cremers, D.: Bregman proximal mappings and Bregman-Moreau envelopes under relative prox-regularity. *J. Optim. Theory Appl.* **184**(3), 724–761 (2020)
37. Lyashko, S.I., Semenov, V.V., Voitova, T.A.: Low-cost modification of Korpelevich's methods for monotone equilibrium problems. *Cybernet. Systems Anal.* **47**(4), 631–639 (2011)
38. Malitsky, Y.: Projected reflected gradient methods for monotone variational inequalities. *SIAM J. Optim.* **25**, 502–520 (2015)
39. Malitsky, Y.: Proximal extrapolated gradient methods for variational inequalities. *Optim. Methods Softw.* **33**(1), 140–164 (2018)
40. Malitsky, Y.: Golden ratio algorithms for variational inequalities. *Math. Program.* **184**(1), 383–410 (2020)
41. Malitsky, Y., Mishchenko, K.: Adaptive gradient descent without descent. In: *Proceedings of the 37th International Conference on Machine Learning*. JMLR.org (2020)
42. Malitsky, Y., Tam, M.K.: A forward-backward splitting method for monotone inclusions without cocoercivity. *SIAM J. Optim.* **30**(2), 1451–1472 (2020)
43. Marcotte, P., Wu, J.H.: On the convergence of projection methods: Application to the decomposition of affine variational inequalities. *J. Optim. Theory Appl.* **85**(2), 347–362 (1995)
44. Nesterov, Y.: Dual extrapolation and its applications to solving variational inequalities and related problems. *Math. Program.* **109**(2), 319–344 (2007)
45. Nomirovskii, D.A., Rublyov, B.V., Semenov, V.V.: Convergence of two-stage method with Bregman divergence for solving variational inequalities. *Cybernet. Syst. Anal.* **55**(3), 359–368 (2019)
46. Pang, J.S., Chan, D.: Iterative methods for variational and complementarity problems. *Math. Program.* **24**, 284–313 (1982)
47. Popov, L.D.: A modification of the Arrow-Hurwicz method for search of saddle points. *Math. Notes Acad. Sci. USSR* **28**(5), 845–848 (1980)
48. Semenov, V.V., Denisov, S.V., Kravets, A.V.: Adaptive two-stage Bregman method for variational inequalities. *Cybernet. Syst. Anal.* **57**(6), 959–967 (2021)
49. Teboulle, M.: A simplified view of first order methods for optimization. *Math. Program.* **170**(1), 67–96 (2018)
50. Wang, W., Carreira-Perpiñán, M.A.: Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application (2013). [arXiv:1309.1541](https://arxiv.org/abs/1309.1541)