Spectral properties of Barzilai-Borwein rules in solving singly linearly constrained optimization problems subject to lower and upper bounds

Serena Crisci[a,d], Federica Porta[b,d,*], Valeria Ruggiero[a,d], Luca Zanni[b,d]

[a] *Department of Mathematics and Computer Science, University of Ferrara, via Machiavelli 30, I-44121 Ferrara, Italy*
[b] *Department of Physics, Informatics and Mathematics, University of Modena and Reggio Emilia, via Campi 213/B, I-41125, Modena, Italy*
[c] *Department of Physics, Informatics and Mathematics, University of Modena and Reggio Emilia, via Campi 213/B, I-41125, Modena, Italy*
[d] *Member of the INdAM Research group GNCS*

Abstract

In 1988, Barzilai and Borwein published a pioneering paper which opened the way to inexpensively accelerate first-order methods. More in detail, in the framework of unconstrained optimization, Barzilai and Borwein developed two strategies to select the steplength in gradient descent methods with the aim of encoding some second-order information of the problem without computing and/or employing the Hessian matrix of the objective function.

Starting from these ideas, several efficient steplength techniques have been suggested in the last decades in order to make gradient descent methods more and more appealing also for problems which handle large-scale data and require real-time solutions. Typically, these new steplength selection rules have been tuned in the quadratic unconstrained framework for sweeping the spectrum of the inverse of the Hessian matrix, and then applied also to non-quadratic constrained problems, without any substantial modification, by showing to be very effective anyway.

In this paper we deeply analyse how, in quadratic and non-quadratic minimization problems, the presence of a feasible region, expressed by a single linear equality constraint together with lower and upper bounds, influences the spectral properties of the original Barzilai-Borwein (BB) rules, generalizing recent results provided for box-constrained quadratic problems. This analysis gives rise to modified BB approaches able not only to capture second-order information but also to exploit the nature of the feasible region. We show the benefits gained by the new steplength rules on a set of test problems arising also from machine learning and image processing applications.

Keywords: Singly Linearly and bound constrained optimization, gradient projection methods, steplength rules, Hessian spectral properties
2010 MSC: 65K05, 90C25, 90C30, 90C06

1. Introduction

In this paper, we are interested in solving the following singly linearly equality constrained optimization problem subject to lower and upper bounds

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} f(x) \\ &\text{subject to } \ \ell \le x \le u \quad v^T x = e \ , \end{aligned} \tag{1}$$

*Corresponding author
Email addresses: serena.crisci@unife.it (Serena Crisci), federica.porta@unimore.it (Federica Porta), valeria.ruggiero@unife.it (Valeria Ruggiero), luca.zanni@unimore.it (Luca Zanni)

where ℓ, u and v are vectors of $\mathbb{R}^n$ and e is a scalar. We assume that the feasible region $\Omega = \{x \in \mathbb{R}^n \ : \ \ell \leq x \leq u \quad v^T x = e\}$ is not empty and the function f is continuously differentiable. We refer to (1) as the general SLB problem. The study of this minimization model is quite relevant since it allows to formalize real-life applications in different areas, such as imaging, signal processing, machine learning and portfolio optimization (see for example [4, 5, 34, 45, 41]). Since a common feature of these applications lies in their large-scale, among all the iterative schemes which can be selected to solve the corresponding optimization problem, the class of gradient projection methods is very attractive thanks to a simple implementation and a low computational cost per iteration. In this work, we consider the so-called gradient projection (GP) method along the feasible direction [6, Chapter 2], whose standard iteration can be written as

$$\begin{aligned} d^{(k)} &= \Pi_\Omega \left(x^{(k)} - \alpha_k \nabla f(x^{(k)}) \right) - x^{(k)}, \\ x^{(k+1)} &= x^{(k)} + \nu_k d^{(k)}, \end{aligned} \tag{2}$$

where $\Pi_\Omega(\cdot)$ denotes the Euclidean projection onto the constraints of (1), α_k is a positive parameter controlling the step along the negative gradient and $\nu_k \in (0, 1]$ is a linesearch parameter ensuring a sufficient decrease of the objective function along the direction $d^{(k)}$, e.g. by means of an Armijo rule [6] or its non-monotone version [30]. Despite the merits previously recalled, the GP method (2) can show a poor convergence rate especially when high accurate solutions are required. A possibility to overcome this difficulty consists in properly selecting the steplength α_k, which simply needs to belong to a compact set $[\alpha_{\min}, \alpha_{\max}]$, $0 < \alpha_{\min} \leq \alpha_{\max}$, in order to guarantee the convergence of the iterative scheme [7]. The literature of the last decades provides many attempts to exploit this freedom of choice for α_k with the aim of accelerating the convergence of the gradient methods. The paper [2] that firstly suggests the key idea to fully take advantage of the presence of the steplength α_k has been published in 1988 by Barzilai and Borwein (BB), in the framework of unconstrained optimization. In that paper, the authors force quasi-Newton properties on the diagonal matrix $(\alpha_k I_n)^{-1}$, where I_n is the identity matrix of order n, for approximating the Hessian matrix $\nabla^2 f(x^{(k)})$. More in detail, the steplength updating rules developed by Barzilai and Borwein have to satisfy the following secant conditions

$$\alpha_k^{\mathrm{BB1}} = \arg\min_{\alpha \in \mathbb{R}} \|\alpha^{-1} s^{(k-1)} - y^{(k-1)}\|, \qquad \alpha_k^{\mathrm{BB2}} = \arg\min_{\alpha \in \mathbb{R}} \|s^{(k-1)} - \alpha y^{(k-1)}\|, \tag{3}$$

where $s^{(k-1)} = x^{(k)} - x^{(k-1)}$ and $y^{(k-1)} = \nabla f(x^{(k)}) - \nabla f(x^{(k-1)})$. The resulting values become

$$\alpha_k^{\mathrm{BB1}} = \frac{{s^{(k-1)}}^T s^{(k-1)}}{{s^{(k-1)}}^T y^{(k-1)}} \qquad ; \qquad \alpha_k^{\mathrm{BB2}} = \frac{{s^{(k-1)}}^T y^{(k-1)}}{{y^{(k-1)}}^T y^{(k-1)}}. \tag{4}$$

We observe that the BB rules (4) are well defined provided that the curvature condition ${s^{(k-1)}}^T y^{(k-1)} > 0$ is satisfied. For this reason, the assumptions of the lemmas and theorems we prove in the following are imposed to ensure the validity of the curvature condition and, hence, the well definiteness of the BB strategies. In particular, when the objective function is strongly convex, the curvature inequality is satisfied for any given points $x^{(k-1)}$ and $x^{(k)}$ [40]. However, the BB schemes are often employed in practice even when weaker assumptions hold. In these cases, negative values of α_k^{BB1} and α_k^{BB2} are properly substituted by an emergency steplength belonging to the interval $[\alpha_{min}, \alpha_{max}]$.

We recall that, in the case of quadratic objective function with symmetric and positive definite Hessian matrix A, the BB rules (4) provide values belonging to the spectrum of the inverse of A, since they obey to the following property

$$\frac{1}{\lambda_{\max}(A)} \leq \alpha_k^{\mathrm{BB2}} \leq \alpha_k^{\mathrm{BB1}} \leq \frac{1}{\lambda_{\min}(A)}, \tag{5}$$

where $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the minimum and the maximum eigenvalues of A, respectively. We briefly remark that the second inequality in (5) follows from the definition of α_k^{BB1} and α_k^{BB2} and the Cauchy-Schwarz inequality [43, Lemma 2.1].

Starting from the inspiring work of Barzilai and Borwein, many steplength updating strategies have been devised to realize more and more effective gradient methods for unconstrained minimization problems (see, for example, [17, 19, 20, 28, 29, 31, 47]). One of the most competitive ideas among the previously recalled strategies is represented by the approach developed in [47], where the authors suggested gradient methods which adaptively alternate small and large steplengths during the iterations. This alternation idea has been successfully exploited in both [28] and [10]. We report here the steplength selection procedure suggested in [10], which alternates the BB rules as follows

$$\alpha_k^{\mathrm{VABB_{min}}} = \begin{cases} \min\{\alpha_j^{\mathrm{BB2}} \ : \ j = \max\{1, k-m_a\}, \ldots, k\}, & \text{if } \frac{\alpha_k^{\mathrm{BB2}}}{\alpha_k^{\mathrm{BB1}}} < \tau_k, \\ \alpha_k^{\mathrm{BB1}} & \text{otherwise,} \end{cases} \tag{6}$$

where m_a is a nonnegative integer and τ_k is updated as

$$\tau_{k+1} = \begin{cases} \tau_k/\zeta & \text{if } \frac{\alpha_k^{\mathrm{BB2}}}{\alpha_k^{\mathrm{BB1}}} < \tau_k, \\ \tau_k \cdot \zeta & \text{otherwise,} \end{cases}$$

with $\zeta > 1$. By selecting $\tau_k = \tau > 0$ at every iteration in (6), the alternating strategy proposed in [28] can be recovered and the corresponding steplength will be hereafter denoted by $\alpha_k^{\mathrm{ABB_{min}}}$.

In the unconstrained quadratic case, the efficiency of a steplength rule within a gradient method is essentially related to its ability to sweep, in a suitable way, the spectrum of the inverse of the underlying Hessian. This understanding, firstly highlighted in [25], has given rise to interesting spectral analysis of the BB-type methods [26, 22], for both quadratic and non-quadratic optimization problems, and has been crucial for devising effective steplength strategies [47, 28, 21], such as the above adaptive rules.

The good results gained in solving unconstrained minimization problems by combining the gradient methods with the BB-like rules encouraged the researchers to exploit these techniques also in gradient projection schemes for constrained problems, by obtaining a great success in different fields [7, 3, 37, 43, 48]. Nevertheless, in these gradient projection approaches, the original BB strategies, and the ones built from them, have been employed without any modification to take into account also the feasible set. Only very recently [16], a spectral analysis of the BB steplength rules in gradient projection methods for box-constrained strictly convex quadratic problems has been developed. By still denoting with ℓ and u the vectors defining the box constraints, the authors introduced the following set of indices

$$\begin{aligned} &\mathcal{I}_{k-1} = \mathcal{N} - \mathcal{J}_{k-1}, \quad \mathcal{N} = \{1, ..., n\}, \\ &\mathcal{J}_{k-1} = \{i \in \mathcal{N} : (x_i^{(k-1)} = \ell_i \wedge x_i^{(k)} = \ell_i) \vee (x_i^{(k-1)} = u_i \wedge x_i^{(k)} = u_i)\}, \end{aligned} \tag{7}$$

and observed that

$$\alpha_k^{\mathrm{BB1}} = \frac{\|s^{(k-1)}\|^2}{{s^{(k-1)}}^T y^{(k-1)}} = \frac{\|s_{\mathcal{I}_{k-1}}^{(k-1)}\|^2}{{s_{\mathcal{I}_{k-1}}^{(k-1)}}^T y_{\mathcal{I}_{k-1}}^{(k-1)}}, \tag{8}$$

where $s_{\mathcal{I}_{k-1}}^{(k-1)}$ and $y_{\mathcal{I}_{k-1}}^{(k-1)}$ represent, respectively, the subvectors of $s^{(k-1)}$ and $y^{(k-1)}$ whose components are indexed in $\mathcal{I}_{k-1}$. Moreover, in view of this consideration and since α_k^{BB2} does not share a similar property, they modified the second BB strategy in the following way

$$\alpha_k^{\text{BOX-BB2}} = \frac{{s_{\mathcal{I}_{k-1}}^{(k-1)}}^T y_{\mathcal{I}_{k-1}}^{(k-1)}}{\|y_{\mathcal{I}_{k-1}}^{(k-1)}\|^2}, \tag{9}$$

and refined conditions (5) as

$$\frac{1}{\lambda_{\max}(A_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}})} \leq \alpha_k^{\text{BOX-BB2}} \leq \alpha_k^{\mathrm{BB1}} \leq \frac{1}{\lambda_{\min}(A_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}})}, \tag{10}$$

by denoting with $A_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}}$ the submatrix of the Hessian of the objective function given by the intersection of the rows and the columns with indices in $\mathcal{I}_{k-1}$. In other words, $1/\alpha_k^{\mathrm{BB1}}$ and $1/\alpha_k^{\text{BOX-BB2}}$ provide

some information about the spectrum of the Hessian submatrix whose rows and columns are indexed in $\mathcal{I}_{k-1}$.

The goal of this paper is to continue the analysis carried out in [16] and to understand how the presence of a single linear equality constraint together with lower and upper bounds can modify the spectral properties of the BB steplength selection rules. Particularly, in the case of quadratic objective function, we investigate how to generalize inequalities (5) and (10) in order to delineate conditions more faithful to the special feasible region of problem (1). As a consequence, we suggested a new version of the BB2 scheme which generalizes (9). Furthermore, in the general non-quadratic case, we study the spectral properties of the considered BB rules providing their interpretation in terms of the Hessian matrices evaluated at the iterates of the gradient projection scheme.

The paper is organized as follows. In Section 2, we develop the spectral analysis of the BB approaches in presence of a strictly convex quadratic SLB problem. The corresponding investigation in the more general non-quadratic framework can be appreciated in Section 3. In Section 4, a generalization of the considered steplength rules to the case of variable metric gradient projection methods is proposed. Section 5 is devoted to the results of the numerical experiments performed on several datasets, concerning both quadratic and non-quadratic problems. The conclusions are drawn in Section 6.

Notation. In the following, we denote by $\mathcal{O}_{r,s}$ the $r \times s$ null matrix (r, s positive integer scalars), by I_r the identity matrix of order r and by $I_{\mathcal{C}}$ the identity matrix of order $\sharp\mathcal{C}$. Moreover, $x_{\mathcal{C}} \in \mathbb{R}^{\sharp\mathcal{C}}$ stands for the subvector of x with entries indexed in $\mathcal{C}$.

2. The quadratic case

We start our analysis from the easier case of a quadratic objective function: the optimization problem we consider in this section has the form

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} f(x) \equiv \frac{1}{2}x^T Ax - b^T x + c \\ &\text{subject to} \quad \ell \le x \le u \quad v^T x = e, \end{aligned} \tag{11}$$

where A is a symmetric positive definite matrix of order n, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Let denote $g(x) = \nabla f(x) = Ax - b$ and by x^* the minimizer of the constrained problem (11). Let $\mathcal{J}^*$ be the set of indices in $\mathcal{N} = \{1, ..., n\}$ of the *active box constraints* at x^*, that is $x_i^* = \ell_i$ or $x_i^* = u_i$, for all $i \in \mathcal{J}^*$, and $\mathcal{I}^* = \mathcal{N}\backslash\mathcal{J}^*$ be the complement of $\mathcal{J}^*$ in $\mathcal{N}$, with cardinality $m = \sharp\mathcal{I}^*$, $0 \le m \le n$. By neglecting the special cases $m = 0$ and $m = n$, we assume that $\mathcal{J}^* \neq \emptyset$, $\mathcal{J}^* \neq \mathcal{N}$, and that v and the columns of the identity matrix of order n with indices in $\mathcal{J}^*$ are linearly independent. Consequently the entries of v corresponding to $\mathcal{I}^*$ are not all equal to zero. For the sake of simplicity, we assume that the rows/columns of A and the entries of any vector are reordered so that $\mathcal{I}^*$ is related to the first m indices and $\mathcal{J}^*$ contains the last $n - m$ indices.

Moreover, we recall that, in view of the necessary and sufficient Karush-Kuhn-Tucker (KKT) conditions, a feasible x^* is a minimizer if and only if there exist $\psi^* \in \mathbb{R}$ and $\mu^*, \nu^* \in \mathbb{R}^n$ such that

$$\begin{cases} g(x^*) - \psi^* v - \mu^* + \nu^* = 0, \\ v^T x^* = e, \\ \mu^* \cdot (x^* - \ell) = 0, \qquad x^* - \ell \ge 0, \qquad \mu^* \ge 0, \\ \nu^* \cdot (u - x^*) = 0, \qquad u - x^* \ge 0, \qquad \nu^* \ge 0, \end{cases} \tag{12}$$

where the products between vectors have to be intended component–wise.

In the following we will show the relation between the BB steplength values and the spectrum of special matrices obtained as restriction of the matrix A to subspaces depending on the constraints which become active during the iterative process. To this end, we first introduce the restriction of A to the tangent space of the active constraints at the solution and then we study the properties of the approximations of this matrix available in the iterations of the gradient projection schemes.

2.1. The Hessian matrix restricted to the tangent space of the active constraints at the solution

The tangent space of the active constraints at x^* is defined as

$$\Omega^* = \left\{ x \in \mathbb{R}^n \ : \ \begin{bmatrix} \mathcal{O}_{n-m,m} & I_{\mathcal{J}^*} \\ v^T & \end{bmatrix} x = \mathcal{O}_{n-m+1,1} \right\} = \text{range}\left(\begin{bmatrix} I_{\mathcal{I}^*} \\ \mathcal{O}_{n-m,m} \end{bmatrix} \right) \ \cap \ \text{null}(v^T).$$

Taking into account that the dimension of Ω^* is $(m-1)$, we introduce the matrix $A^* \in \mathbb{R}^{(m-1)\times(m-1)}$ defined as

$$A^* = {\tilde{U}^*}^T A \tilde{U}^*, \tag{13}$$

where $\tilde{U}^*$ denotes an $n \times (m-1)$ matrix whose columns are an orthonormal basis of Ω^*. We call A^* the Hessian matrix restricted to the tangent space of the active constraints at the solution, hereafter named *restricted Hessian matrix*. In order to characterize the symmetric positive definite matrix A^*, we collect some useful results on projection matrices in Lemma 2.1 (see [33]).

Lemma 2.1. *Let u be a non-zero vector in $\mathbb{R}^m$ and define the matrices $V = \frac{uu^T}{u^Tu}$ and $P = I_m - V$. It holds that*

(a) the matrix V is the orthogonal projection onto the subspace $\mathcal{V} = range(u)$; the matrix P is the orthogonal projection onto $\mathcal{V}^\perp = range(u)^\perp = null(u^T)$;

(b) $V = V^T = V^2$ and $P = P^T = P^2$;

(c) $range(V) = null(P) = \mathcal{V}$ and $null(V) = range(P) = \mathcal{V}^\perp$;

(d) the spectral decomposition of V is $V = W \begin{bmatrix} \mathcal{O}_{m-1,m-1} & \mathcal{O}_{m-1,1} \\ \mathcal{O}_{1,m-1} & 1 \end{bmatrix} W^T$, where $W = \begin{bmatrix} \tilde{W} & w \end{bmatrix}$ is an orthogonal matrix of order m, with $\tilde{W} \in \mathbb{R}^{m\times(m-1)}$, $w \in \mathbb{R}^m$; the eigenvalues of V are either 1 or 0; $range(V) = range(w)$ is the one-dimensional eigenspace associated to the eigenvalue 1 and $null(V) = range(\tilde{W})$ is the eigenspace of dimension $m-1$ associated to the eigenvalue 0;

(e) $P = W \begin{bmatrix} I_{m-1} & \mathcal{O}_{m-1,1} \\ \mathcal{O}_{1,m-1} & 0 \end{bmatrix} W^T$ is the spectral decomposition of P; the eigenvalues of P are either 1 or 0; $range(P) = range(\tilde{W})$ is the eigenspace of dimension $m-1$ associated to the eigenvalue 1 and $null(P) = range(w)$ is the one-dimensional eigenspace associated to the eigenvalue 0;

(f) $P = \tilde{W}\tilde{W}^T$ and $V = ww^T$, with $\tilde{W}^T\tilde{W} = I_{m-1}$ and $w^Tw = 1$, $w = \frac{u}{\|u\|}$; furthermore $\tilde{W}$ is an orthonormal basis of $range(P) = \mathcal{V}^\perp = null(u^T)$.

By using the notation $v^T = \begin{bmatrix} v_{\mathcal{I}^*}^T & v_{\mathcal{J}^*}^T \end{bmatrix}$ and by applying Lemma 2.1 with $u = v_{\mathcal{I}^*}$ it is possible to construct the matrix $\tilde{U}^*$ used in the definition of A^*. Indeed, if we denote by P^* the orthogonal projection onto $\text{null}(v_{\mathcal{I}^*}^T)$, thanks to part (f) of Lemma 2.1, there exists a matrix $\tilde{W}^* \in \mathbb{R}^{m\times(m-1)}$ whose columns are an orthonormal basis for $\text{null}(v_{\mathcal{I}^*}^T)$. Consequently, the matrix $\tilde{U}^* = \begin{bmatrix} \tilde{W}^* \\ \mathcal{O}_{n-m,m-1} \end{bmatrix}$ provides an orthonormal basis for Ω^*, since any n-vector $x \in \Omega^*$ can be expressed as $x^T = \begin{bmatrix} x_{\mathcal{I}^*}^T & x_{\mathcal{J}^*}^T \end{bmatrix}$ with $x_{\mathcal{I}^*} \in \text{null}(v_{\mathcal{I}^*}^T)$ and $x_{\mathcal{J}^*} = \mathcal{O}_{n-m,1}$.

In the next subsection we will show that, when a gradient projection method is applied for solving problem (11), the spectrum of A^* plays a crucial role in the analysis of the steplength rules.

2.2. The spectral properties of the BB rules in terms of the approximated restricted Hessian matrices

Of course, at the beginning of the iterative process generated by a gradient projection method, the solution x^* of (11) is not known and, for our study, we need to focus on a sequence of matrices which approximate A^* during the iterations. First of all, we provide a way to realize such a sequence of approximation matrices and we underline the relationship between their spectra and the BB steplength selection rules. To achieve this goal, we consider the set of indices introduced in (7). Also in this case, for

the sake of simplicity, we assume that the rows/columns of A and the entries of any vector are reordered so that $\mathcal{I}_{k-1}$ is related to the first $m_k = \sharp\mathcal{I}_{k-1}$ indices and $\mathcal{J}_{k-1}$ contains the last $n - m_k$ indices.

If we consider the orthogonal projection P_{k-1} onto null$(v_{\mathcal{I}_{k-1}}^T)$,

$$P_{k-1} = I_{\mathcal{I}_{k-1}} - \frac{1}{v_{\mathcal{I}_{k-1}}^T v_{\mathcal{I}_{k-1}}} v_{\mathcal{I}_{k-1}} v_{\mathcal{I}_{k-1}}^T, \tag{14}$$

from part (f) of the Lemma 2.1, we have that there exists a $m_k \times (m_k - 1)$ matrix $\tilde{W}_{k-1}$ with orthonormal columns such that

$$P_{k-1} = \tilde{W}_{k-1} \tilde{W}_{k-1}^T, \tag{15}$$

and the $n \times (m_k - 1)$ matrix

$$\tilde{U}_{k-1} = \begin{bmatrix} \tilde{W}_{k-1} \\ \mathcal{O}_{n-m_k, m_k-1} \end{bmatrix} \tag{16}$$

is an orthonormal basis for the subspace

$$\Omega_{k-1} = \left\{ x \in \mathbb{R}^n \ : \ \begin{bmatrix} \mathcal{O}_{n-m_k, m_k} & I_{\mathcal{J}_{k-1}} \\ v^T & \end{bmatrix} x = \mathcal{O}_{n-m_k+1,1} \right\}. \tag{17}$$

Therefore, the symmetric positive definite matrix $\tilde{U}_{k-1}^T A \tilde{U}_{k-1}$ represents an approximation of the matrix (13) at the iteration k.

Now, by using the notation $g^{(k)} = g(x^{(k)})$, we consider $y^{(k-1)} = g^{(k)} - g^{(k-1)}$ and we introduce the following vector:

$$t^{(k-1)} = g^{(k)} - \psi_k v - (g^{(k-1)} - \psi_{k-1} v) = y^{(k-1)} - (\psi_k - \psi_{k-1}) v, \tag{18}$$

where ψ_{k-1} and ψ_k are approximations of the equality constraint multiplier ψ^* computed at the iterations $k-1$ and k, respectively, (see the conditions (12)):

$$\psi_{k-1} = \frac{v_{\mathcal{I}_{k-1}}^T g_{\mathcal{I}_{k-1}}^{(k-1)}}{v_{\mathcal{I}_{k-1}}^T v_{\mathcal{I}_{k-1}}}, \qquad \psi_k = \frac{v_{\mathcal{I}_{k-1}}^T g_{\mathcal{I}_{k-1}}^{(k)}}{v_{\mathcal{I}_{k-1}}^T v_{\mathcal{I}_{k-1}}}. \tag{19}$$

Note that quantities similar to $t^{(k-1)}$ were considered also in the framework of interior point methods [32].

The vector (18) can be written as $t^{(k-1)} = \begin{bmatrix} t_{\mathcal{I}_{k-1}}^{(k-1)} \\ t_{\mathcal{J}_{k-1}}^{(k-1)} \end{bmatrix}$ and the following equalities hold:

$$t_{\mathcal{I}_{k-1}}^{(k-1)} = P_{k-1} y_{\mathcal{I}_{k-1}}^{(k-1)} = \tilde{W}_{k-1} \tilde{W}_{k-1}^T y_{\mathcal{I}_{k-1}}^{(k-1)}. \tag{20}$$

The next Lemma is useful to understand the role of $t_{\mathcal{I}_{k-1}}^{(k-1)}$ in the definitions of the BB steplength rules.

Lemma 2.2. *Given $s^{(k-1)} = x^{(k)} - x^{(k-1)}$ and $t^{(k-1)}$ as in (18), it holds that*

(a) ${s^{(k-1)}}^T v = 0$;

(b) ${s_{\mathcal{I}_{k-1}}^{(k-1)}}^T v_{\mathcal{I}_{k-1}} = 0$;

(c) ${s^{(k-1)}}^T y^{(k-1)} = {s^{(k-1)}}^T t^{(k-1)} = {s_{\mathcal{I}_{k-1}}^{(k-1)}}^T t_{\mathcal{I}_{k-1}}^{(k-1)}$.

Proof. (a) ${s^{(k-1)}}^T v = 0$ since both $x^{(k-1)}$ and $x^{(k)}$ satisfy the equality constraint $v^T x = e$.
(b) We show that ${s_{\mathcal{I}_{k-1}}^{(k-1)}}^T v_{\mathcal{I}_{k-1}} = 0$. Indeed,

$$0 = {s^{(k-1)}}^T v = \sum_{i \in \mathcal{I}_{k-1}} {s_i^{(k-1)}}^T v_i + \sum_{i \in \mathcal{J}_{k-1}} {s_i^{(k-1)}}^T v_i = \sum_{i \in \mathcal{I}_{k-1}} {s_i^{(k-1)}}^T v_i,$$

where the last equality holds since $s^{(k-1)}_{\mathcal{J}_{k-1}} = \mathcal{O}_{n-m_k,1}$ from the definition of $\mathcal{J}_{k-1}$.
(c) From parts (a), we have that

$$s^{(k-1)^T} t^{(k-1)} = s^{(k-1)^T} \left(y^{(k-1)} - (\psi_k - \psi_{k-1}) v \right) = s^{(k-1)^T} y^{(k-1)}, \tag{21}$$

and, since $s^{(k-1)}_{\mathcal{J}_{k-1}} = \mathcal{O}_{n-m_k,1}$, the last equality follows easily. □

Lemma 2.2 allows us to state that the classical formulation (4) of the first BB rule provides a steplength depending only on the indices belonging to the set $\mathcal{I}_{k-1}$:

$$\alpha_k^{\text{BB1}} = \frac{s^{(k-1)^T} s^{(k-1)}}{s^{(k-1)^T} y^{(k-1)}} = \frac{s^{(k-1)^T}_{\mathcal{I}_{k-1}} s^{(k-1)}_{\mathcal{I}_{k-1}}}{s^{(k-1)^T}_{\mathcal{I}_{k-1}} t^{(k-1)}_{\mathcal{I}_{k-1}}}. \tag{22}$$

In some sense, the rule α_k^{BB1} computes the steplength by capturing information from the current inactive constraints, discarding the effect of the constraints that remain active in the last two iterations. The original BB2 rule does not fulfill a similar property, due to the special form of its denominator. This remark suggests to study the properties of the following modified BB2 rule:

$$\alpha_k^{\text{EQ-BB2}} = \frac{s^{(k-1)^T} y^{(k-1)}}{t^{(k-1)^T}_{\mathcal{I}_{k-1}} t^{(k-1)}_{\mathcal{I}_{k-1}}}. \tag{23}$$

Theorem 2.1 shows that the reciprocals of α_k^{BB1} and $\alpha_k^{\text{EQ-BB2}}$ give spectral information about the matrix $\tilde{U}_{k-1}^T A \tilde{U}_{k-1}$.

Theorem 2.1. *Under the assumption that the matrix A is symmetric and positive definite, we have*

$$\lambda_{\min}(\tilde{U}_{k-1}^T A \tilde{U}_{k-1}) \leq 1/\alpha_k^{\text{BB1}} \leq \lambda_{\max}(\tilde{U}_{k-1}^T A \tilde{U}_{k-1}), \tag{24}$$

$$\lambda_{\min}(\tilde{U}_{k-1}^T A \tilde{U}_{k-1}) \leq 1/\alpha_k^{\text{EQ-BB2}} \leq \lambda_{\max}(\tilde{U}_{k-1}^T A \tilde{U}_{k-1}). \tag{25}$$

Proof. In the following, we drop for simplicity the iteration counter $k-1$ from $\mathcal{I}_{k-1}$ and $\mathcal{J}_{k-1}$. In view of the gradient projection iteration (2), we have that the entries of the iterate $x^{(k)}$ are

$$x_i^{(k)} = \begin{cases} x_i^{(k-1)} + \nu_{k-1}(r_i^{(k-1)} - x_i^{(k-1)}) & \text{for } i \in \mathcal{I}, \\ x_i^{(k-1)} & \text{for } i \in \mathcal{J}, \end{cases} \tag{26}$$

where $r_i^{(k-1)} = (\Pi_\Omega(x^{(k-1)} - \alpha_{k-1} g^{(k-1)}))_i, i \in \mathcal{I}$. The vector $s^{(k-1)}$ can be partitioned into two subvectors as follows

$$s^{(k-1)} = \begin{pmatrix} s_{\mathcal{I}}^{(k-1)} \\ s_{\mathcal{J}}^{(k-1)} \end{pmatrix} = \begin{pmatrix} \nu_{k-1}(r^{(k-1)} - x_{\mathcal{I}}^{(k-1)}) \\ \mathcal{O}_{\mathcal{J},1} \end{pmatrix}. \tag{27}$$

Any entry $g_i^{(k)}, i = 1, \ldots, n$, of the gradient $g^{(k)}$ has the following expression:

$$\begin{aligned} g_i^{(k)} &= \sum_{j=1}^{n} a_{ij} x_j^{(k)} - b_i \\ &= \sum_{j \in \mathcal{I}} a_{ij}(x_j^{(k-1)} + \nu_{k-1}(r_j^{(k-1)} - x_j^{(k-1)})) + \sum_{j \in \mathcal{J}} a_{ij} x_j^{(k-1)} - b_i \\ &= g_i^{(k-1)} + \nu_{k-1} \sum_{j \in \mathcal{I}} a_{ij}(r_j^{(k-1)} - x_j^{(k-1)}). \end{aligned}$$

Consequently, from (27), we can write

$$y^{(k-1)} = \begin{pmatrix} y_{\mathcal{I}}^{(k-1)} \\ y_{\mathcal{J}}^{(k-1)} \end{pmatrix} = \begin{pmatrix} A_{\mathcal{I},\mathcal{I}} s_{\mathcal{I}}^{(k-1)} \\ A_{\mathcal{J},\mathcal{I}} s_{\mathcal{I}}^{(k-1)} \end{pmatrix}. \tag{28}$$

Furthermore, from part (b) of Lemma 2.2, it follows that

$$P_{k-1} s_{\mathcal{I}}^{(k-1)} = \left(I_{\mathcal{I}} - \frac{1}{v_{\mathcal{I}}^T v_{\mathcal{I}}} v_{\mathcal{I}} v_{\mathcal{I}}^T \right) s_{\mathcal{I}}^{(k-1)} = s_{\mathcal{I}}^{(k-1)}. \tag{29}$$

Hence, from (29) and (15) we observe that

$$s^{(k-1)^T} s^{(k-1)} = s_{\mathcal{I}}^{(k-1)^T} s_{\mathcal{I}}^{(k-1)} = s_{\mathcal{I}}^{(k-1)^T} P_{k-1} s_{\mathcal{I}}^{(k-1)} = \|\tilde{W}_{k-1}^T s_{\mathcal{I}}^{(k-1)}\|^2. \tag{30}$$

Moreover, from part (c) of Lemma 2.2, (20), (28), (29), (15) and (16), we obtain

$$\begin{aligned} s^{(k-1)^T} y^{(k-1)} &= s_{\mathcal{I}}^{(k-1)^T} t_{\mathcal{I}}^{(k-1)} = s_{\mathcal{I}}^{(k-1)^T} P_{k-1} y_{\mathcal{I}}^{(k-1)} \\ &= s_{\mathcal{I}}^{(k-1)^T} P_{k-1} A_{\mathcal{I},\mathcal{I}} s_{\mathcal{I}}^{(k-1)} = s_{\mathcal{I}}^{(k-1)^T} P_{k-1} A_{\mathcal{I},\mathcal{I}} P_{k-1} s_{\mathcal{I}}^{(k-1)} \\ &= s_{\mathcal{I}}^{(k-1)^T} \tilde{W}_{k-1} \tilde{W}_{k-1}^T A_{\mathcal{I},\mathcal{I}} \tilde{W}_{k-1} \tilde{W}_{k-1}^T s_{\mathcal{I}}^{(k-1)} \\ &= s_{\mathcal{I}}^{(k-1)^T} \tilde{W}_{k-1} \tilde{U}_{k-1}^T A \tilde{U}_{k-1} \tilde{W}_{k-1}^T s_{\mathcal{I}}^{(k-1)}. \end{aligned} \tag{31}$$

From (31) and (30), we conclude that $1/\alpha_k^{\text{BB1}}$ is the Rayleigh quotient of the matrix $\tilde{U}_{k-1}^T A \tilde{U}_{k-1}$ at the vector $\tilde{W}_{k-1}^T s_{\mathcal{I}}^{(k-1)}$ and the inequality (24) holds.

Furthermore, by proceeding as for the equalities (31), it is immediate to write

$$t_{\mathcal{I}}^{(k-1)} = \tilde{W}_{k-1} \tilde{U}_{k-1}^T A \tilde{U}_{k-1} \tilde{W}_{k-1}^T s_{\mathcal{I}}^{(k-1)}. \tag{32}$$

As a consequence, since $\tilde{W}_{k-1}^T \tilde{W}_{k-1} = I_{m-1}$, we obtain

$$t_{\mathcal{I}}^{(k-1)^T} t_{\mathcal{I}}^{(k-1)} = s_{\mathcal{I}}^{(k-1)^T} \tilde{W}_{k-1} (\tilde{U}_{k-1}^T A \tilde{U}_{k-1})^2 \tilde{W}_{k-1}^T s_{\mathcal{I}}^{(k-1)}. \tag{33}$$

Since $\tilde{U}_{k-1}^T A \tilde{U}_{k-1}$ is a symmetric positive definite matrix, we can introduce the vector

$$z^{(k-1)} = (\tilde{U}_{k-1}^T A \tilde{U}_{k-1})^{1/2} \tilde{W}_{k-1}^T s_{\mathcal{I}}^{(k-1)}, \tag{34}$$

so that the scalar product in (33) can be written as

$$t_{\mathcal{I}}^{(k-1)^T} t_{\mathcal{I}}^{(k-1)} = z^{(k-1)^T} \tilde{U}_{k-1}^T A \tilde{U}_{k-1} z^{(k-1)}, \tag{35}$$

and $s^{(k-1)^T} y^{(k-1)} = z^{(k-1)^T} z^{(k-1)}$; thus, $1/\alpha_k^{\text{EQ-BB2}}$ is the Rayleigh quotient of the matrix $\tilde{U}_{k-1}^T A \tilde{U}_{k-1}$ at the vector $z^{(k-1)}$ and the inequality (25) holds. □

The modified BB2 rule (23) not only exhibits the same spectral properties of the BB1 steplength but it also allows to recover a relationship between the two rules analogous to (5) and (10), as the next theorem shows.

Theorem 2.2. *The steplengths α_k^{BB1} and $\alpha_k^{\text{EQ-BB2}}$ satisfy $\alpha_k^{\text{EQ-BB2}} \leq \alpha_k^{\text{BB1}}$.*

Proof. From the Cauchy-Schwarz inequality, it follows that

$$\begin{aligned} \frac{1}{\alpha_k^{\text{BB1}}} &= \frac{s^{(k-1)^T} y^{(k-1)}}{s^{(k-1)^T} s^{(k-1)}} = \frac{s_{\mathcal{I}_{k-1}}^{(k-1)^T} t_{\mathcal{I}_{k-1}}^{(k-1)}}{s_{\mathcal{I}_{k-1}}^{(k-1)^T} s_{\mathcal{I}_{k-1}}^{(k-1)}} \leq \frac{\|s_{\mathcal{I}_{k-1}}^{(k-1)}\| \|t_{\mathcal{I}_{k-1}}^{(k-1)}\|}{\|s_{\mathcal{I}_{k-1}}^{(k-1)}\|^2} \\ &= \frac{\|t_{\mathcal{I}_{k-1}}^{(k-1)}\| \|t_{\mathcal{I}_{k-1}}^{(k-1)}\|}{\|s_{\mathcal{I}_{k-1}}^{(k-1)}\| \|t_{\mathcal{I}_{k-1}}^{(k-1)}\|} \leq \frac{t_{\mathcal{I}_{k-1}}^{(k-1)^T} t_{\mathcal{I}_{k-1}}^{(k-1)}}{s_{\mathcal{I}_{k-1}}^{(k-1)^T} t_{\mathcal{I}_{k-1}}^{(k-1)}} = \frac{t_{\mathcal{I}_{k-1}}^{(k-1)^T} t_{\mathcal{I}_{k-1}}^{(k-1)}}{s^{(k-1)^T} y^{(k-1)}} = \frac{1}{\alpha_k^{\text{EQ-BB2}}}. \end{aligned}$$

□

The spectral properties described in Theorem 2.1 are useful for relating the steplength to the ability of the gradient projection scheme to annihilate the quantity $g^{(k-1)}_{\mathcal{I}_{k-1}} - \psi_{k-1} v_{\mathcal{I}_{k-1}}$, that is a remarkable skill since at the solution x^* we have $g_{\mathcal{I}^*}(x^*) - \psi^* v_{\mathcal{I}^*} = 0$, as ensured by the KKT conditions (12). In order to explain this fact, by supposing $\mathcal{I}_{k-1} = \mathcal{I}_k$, we observe that

$$\begin{aligned} g^{(k)}_{\mathcal{I}_k} - \psi_k v_{\mathcal{I}_k} &= g^{(k)}_{\mathcal{I}_k} - \frac{v^T_{\mathcal{I}_k} g^{(k)}_{\mathcal{I}_k}}{v^T_{\mathcal{I}_k} v_{\mathcal{I}_k}} v_{\mathcal{I}_k} = P_k g^{(k)}_{\mathcal{I}_k}, \\ g^{(k-1)}_{\mathcal{I}_{k-1}} - \psi_{k-1} v_{\mathcal{I}_{k-1}} &= g^{(k-1)}_{\mathcal{I}_{k-1}} - \frac{v^T_{\mathcal{I}_{k-1}} g^{(k-1)}_{\mathcal{I}_{k-1}}}{v^T_{\mathcal{I}_{k-1}} v_{\mathcal{I}_{k-1}}} v_{\mathcal{I}_{k-1}} = P_{k-1} g^{(k-1)}_{\mathcal{I}_{k-1}}, \end{aligned} \tag{36}$$

and we show in Theorem 2.3 how a gradient projection step affects the vector $P_k g^{(k)}_{\mathcal{I}_k}$.

Theorem 2.3. *Assume that $\mathcal{I}_{k-1} = \mathcal{I}_k$ and $\ell_i < \left(\Pi_\Omega(x^{(k-1)} - \alpha_{k-1} g^{(k-1)})\right)_i < u_i$, $i \in \mathcal{I}_{k-1}$. The following equalities hold:*

$$P_k g^{(k)}_{\mathcal{I}_k} = \left(I_{m_k} - \alpha_{k-1}\nu_{k-1} P_{k-1} A_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}} P_{k-1}\right) P_{k-1} g^{(k-1)}_{\mathcal{I}_{k-1}}, \tag{37}$$

$$\tilde{W}_k \tilde{W}_k^T g^{(k)}_{\mathcal{I}_k} = \tilde{W}_{k-1}\left(I_{m_k-1} - \alpha_{k-1}\nu_{k-1}\tilde{U}^T_{k-1} A \tilde{U}_{k-1}\right)\tilde{W}^T_{k-1} g^{(k-1)}_{\mathcal{I}_{k-1}}, \tag{38}$$

where $\tilde{W}_{k-1} \in \mathbb{R}^{m_k \times (m_k-1)}$ and $\tilde{W}_k \in \mathbb{R}^{m_{k+1} \times (m_{k+1}-1)}$ are matrices with orthonormal columns such that $P_{k-1} = \tilde{W}_{k-1}\tilde{W}^T_{k-1}$ and $P_k = \tilde{W}_k \tilde{W}_k^T$, respectively.

Proof. By using equations (27) and (29) and the hypothesis $\mathcal{I}_{k-1} = \mathcal{I}_k$, we can write

$$\begin{aligned} P_k g^{(k)}_{\mathcal{I}_k} &= P_{k-1}\left(g^{(k-1)}_{\mathcal{I}_{k-1}} + A_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}} s^{(k-1)}_{\mathcal{I}_{k-1}}\right) \\ &= P_{k-1} g^{(k-1)}_{\mathcal{I}_{k-1}} + \nu_{k-1} P_{k-1} A_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}} P_{k-1}\left(r^{(k-1)} - x^{(k-1)}_{\mathcal{I}_{k-1}}\right), \end{aligned} \tag{39}$$

where, as before, $r^{(k-1)} = \left(\Pi_\Omega(x^{(k-1)} - \alpha_{k-1} g^{(k-1)})\right)_{\mathcal{I}_{k-1}}$. Since $x^{(k)}_{\mathcal{J}_{k-1}} = x^{(k-1)}_{\mathcal{J}_{k-1}}$ (from the definition of $\mathcal{J}_{k-1}$) and $\ell_{\mathcal{I}_{k-1}} < r^{(k-1)} < u_{\mathcal{I}_{k-1}}$, the vector $\Pi_\Omega(x^{(k-1)} - \alpha_{k-1} g^{(k-1)})$ can be written as

$$\Pi_\Omega(x^{(k-1)} - \alpha_{k-1} g^{(k-1)}) = \begin{pmatrix} r^{(k-1)} \\ x^{(k-1)}_{\mathcal{J}_{k-1}} \end{pmatrix},$$

where, by using by notation $\tilde{e} = e - v^T_{\mathcal{J}_{k-1}} x^{(k-1)}_{\mathcal{J}_{k-1}}$, $r^{(k-1)}$ solves the following problem

$$r^{(k-1)} = \underset{\left\{r \,:\, v^T_{\mathcal{I}_{k-1}} r = \tilde{e}\right\}}{\arg\min} \frac{1}{2}\|r - (x^{(k-1)} - \alpha_{k-1} g^{(k-1)})_{\mathcal{I}_{k-1}}\|^2. \tag{40}$$

From the KKT conditions related to the problem (40), the vector $r^{(k-1)}$ has the following expression

$$\begin{aligned} r^{(k-1)} &= x^{(k-1)}_{\mathcal{I}_{k-1}} - \alpha_{k-1} g^{(k-1)}_{\mathcal{I}_{k-1}} - \frac{v^T_{\mathcal{I}_{k-1}}(x^{(k-1)}_{\mathcal{I}_{k-1}} - \alpha_{k-1} g^{(k-1)}_{\mathcal{I}_{k-1}})}{v^T_{\mathcal{I}_{k-1}} v_{\mathcal{I}_{k-1}}} v_{\mathcal{I}_{k-1}} + \frac{\tilde{e} v_{\mathcal{I}_{k-1}}}{v^T_{\mathcal{I}_{k-1}} v_{\mathcal{I}_{k-1}}} = \\ &= x^{(k-1)}_{\mathcal{I}_{k-1}} - \alpha_{k-1} g^{(k-1)}_{\mathcal{I}_{k-1}} - \frac{v^T_{\mathcal{I}_{k-1}} x^{(k-1)}_{\mathcal{I}_{k-1}} v_{\mathcal{I}_{k-1}}}{v^T_{\mathcal{I}_{k-1}} v_{\mathcal{I}_{k-1}}} + \frac{\alpha_{k-1} v^T_{\mathcal{I}_{k-1}} g^{(k-1)}_{\mathcal{I}_{k-1}} v_{\mathcal{I}_{k-1}}}{v^T_{\mathcal{I}_{k-1}} v_{\mathcal{I}_{k-1}}} + \\ &\quad + \frac{\tilde{e} v_{\mathcal{I}_{k-1}}}{v^T_{\mathcal{I}_{k-1}} v_{\mathcal{I}_{k-1}}} = x^{(k-1)}_{\mathcal{I}_{k-1}} - \alpha_{k-1} P_{k-1} g^{(k-1)}_{\mathcal{I}_{k-1}}. \end{aligned}$$

From equation (39) and the previous one, the first equality of the thesis is proved. The second one follows easily from the definition of $\tilde{W}_{k-1}$ and $\tilde{W}_k$. □

Thanks to the assumption $\mathcal{I}_k = \mathcal{I}_{k-1}$, we have that $\tilde{W}_{k-1} = \tilde{W}_k$ and, due to the linear independence of their columns, from (38) it follows that

$$\tilde{W}_k^T g_{\mathcal{I}_k}^{(k)} \;=\; \left(I_{m_k-1} - \alpha_{k-1}\nu_{k-1}\tilde{U}_{k-1}^T A\tilde{U}_{k-1}\right)\tilde{W}_{k-1}^T g_{\mathcal{I}_{k-1}}^{(k-1)}. \tag{41}$$

If we denote by $\lambda_1, \ldots, \lambda_{m_k-1}$ and $\xi_1, \ldots, \xi_{m_k-1}$ the eigenvalues and the associated orthonormal eigenvectors of $\tilde{U}_{k-1}^T A\tilde{U}_{k-1}$, we may write $\tilde{W}_k^T g_{\mathcal{I}_k}^{(k)} = \sum_{i=1}^{m_k-1}\gamma_i^{(k)}\xi_i$ and $\tilde{W}_{k-1}^T g_{\mathcal{I}_{k-1}}^{(k-1)} = \sum_{i=1}^{m_k-1}\gamma_i^{(k-1)}\xi_i$. For the eigencomponents $\gamma_i^{(k)}$ the following recurrence formula can be easily derived from (41)

$$\gamma_i^{(k)} \;=\; (1 - \alpha_{k-1}\nu_{k-1}\lambda_i)\,\gamma_i^{(k-1)}, \qquad i = 1, \ldots, m_k - 1. \tag{42}$$

Formula (42) highlights that if α_{k-1} is an accurate approximation of the inverse of an eigenvalue of $\tilde{U}_{k-1}^T A\tilde{U}_{k-1}$, since $\nu_k \in (0, 1]$, a reduction of $\left|\gamma_i^{(k)}\right|$ with respect to $\left|\gamma_i^{(k-1)}\right|$ is obtained. By remembering that $\tilde{W}^{*T} g_{\mathcal{I}^*}(x^*) = 0$, we conclude that the use of a steplength rule providing good approximations of the inverse of the eigenvalues of $\tilde{U}_{k-1}^T A\tilde{U}_{k-1}$ can be a fruitful strategy for accelerating gradient projection methods for problem (11).

In analogy with the box-constrained case [16], the previous theorems suggest that the modified BB2 rule (23) can be exploited within the adaptive strategy (6) in order to design a steplength selection that better sweeps the spectrum of the inverse of $\tilde{U}_{k-1}^T A\tilde{U}_{k-1}$ with respect to the original rule (6). The resulting scheme can be written as

$$\alpha_k^{\text{EQ-VABB}_{\min}} = \begin{cases} \min\{\alpha_j^{\text{EQ-BB2}} \;:\; j = \max\{1, k-m_a\}, \ldots, k\} & \text{if } \dfrac{\alpha_k^{\text{EQ-BB2}}}{\alpha_k^{\text{BB1}}} < \tau_k, \\ \alpha_k^{\text{BB1}} & \text{otherwise,} \end{cases} \tag{43}$$

where m_a and τ_k are defined as in (6). We can guarantee that $1/\alpha_k^{\text{EQ-VABB}_{\min}}$ belongs to the spectrum of $\tilde{U}_{k-1}^T A\tilde{U}_{k-1}$ at any iteration only if $m_a = 0$. Indeed, if $m_a > 0$, inequalities (25) do not hold, in general, for $\alpha_j^{\text{EQ-BB2}}$ with $j = \max\{1, k - m_a\}, \ldots, k - 1$. However, small values for m_a are acceptable since the final active set stabilizes at some point of the iterative process. In the following we will denote by $\alpha_k^{\text{EQ-ABB}_{\min}}$ the version of (43) with $\tau_k = \tau$, $\forall k$, that is the steplength strategy suggested in [28] properly modified to account for the single linear equality constraint and the lower and upper bounds.

Remark 2.1. Before concluding this section, we mention how to deal with a slight modification of the feasible set in (11): a single linear *inequality* constraint instead of a single linear *equality* one. In case of an inequality constraint, if in two successive iterations the linear constraint is active then the BB2 strategy has to be defined according to (23), otherwise it must be fixed according to (9) which only takes into account the presence of the lower and the upper bounds.

3. The non-quadratic case

In this section we come back to the original non-quadratic problem (1) and we define the spectral properties of α_k^{BB1} and $\alpha_k^{\text{EQ-BB2}}$ with respect to the Hessian matrix of the objective function at $x^{(k)}$. The multidimensional variant of Taylor's theorem [40, Theorem 11.1] allows to write the following equation:

$$y^{(k-1)} = \nabla f(x^{(k)}) - \nabla f(x^{(k-1)}) = \int_0^1 \nabla^2 f(x^{(k-1)} + \varrho s^{(k-1)}) s^{(k-1)} d\varrho. \tag{44}$$

From (20) and (44) and by recalling that $s_{\mathcal{J}_{k-1}}^{(k-1)} = 0$, it holds

$$t_{\mathcal{I}_{k-1}}^{(k-1)} = [P_{k-1} \;\; \mathcal{O}_{m_k, n-m_k}]\, y^{(k-1)} = \int_0^1 P_{k-1}\nabla^2 f(x^{(k-1)} + \varrho s^{(k-1)})_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}} s_{\mathcal{I}_{k-1}}^{(k-1)} d\varrho,$$

and, from part (c) of Lemma 2.2 and the equality $s^{(k-1)}_{\mathcal{I}_{k-1}} = P_{k-1}s^{(k-1)}_{\mathcal{I}_{k-1}}$, we have

$$\begin{aligned} {s^{(k-1)}}^T y^{(k-1)} &= {s^{(k-1)}_{\mathcal{I}_{k-1}}}^T t^{(k-1)}_{\mathcal{I}_{k-1}} \\ &= {s^{(k-1)}_{\mathcal{I}_{k-1}}}^T \int_0^1 \tilde{W}_{k-1}\tilde{W}_{k-1}^T \nabla^2 f(x^{(k-1)} + \varrho s^{(k-1)})_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}} \tilde{W}_{k-1}\tilde{W}_{k-1}^T s^{(k-1)}_{\mathcal{I}_{k-1}} d\varrho \\ &= {s^{(k-1)}_{\mathcal{I}_{k-1}}}^T \tilde{W}_{k-1} \int_0^1 \left(\tilde{U}_{k-1}^T \nabla^2 f(x^{(k-1)} + \varrho s^{(k-1)})\tilde{U}_{k-1}\right) \tilde{W}_{k-1}^T s^{(k-1)}_{\mathcal{I}_{k-1}} d\varrho. \end{aligned}$$

Since ${s^{(k-1)}_{\mathcal{I}_{k-1}}}^T \tilde{W}_{k-1}\tilde{W}_{k-1}^T s^{(k-1)}_{\mathcal{I}_{k-1}} = {s^{(k-1)}}^T s^{(k-1)}$, from the last equality we can conclude that $1/\alpha_k^{\text{EQ-BB1}}$ can be interpreted as a Rayleigh quotient relative to the average matrix $\tilde{U}_{k-1}^T \nabla^2 f(x^{(k-1)} + \varrho s^{(k-1)})\tilde{U}_{k-1}$.

In order to give a similar interpretation for $\alpha_k^{\text{EQ-BB2}}$, we take into account the linear function $\phi\colon \mathbb{R}^{m-1} \to \mathbb{R}^{m-1}$ defined as

$$\phi(x_{\tilde{W}}) = z_{\tilde{W}} = \tilde{W}_{k-1}^T P_{k-1} \nabla f(x_{\mathcal{I}_{k-1}}, x_{\mathcal{J}_{k-1}})_{\mathcal{I}_{k-1}},$$

where $x_{\tilde{W}} \in \mathbb{R}^{m-1}$, $x_{\mathcal{I}_{k-1}} = \tilde{W}_{k-1}x_{\tilde{W}}$ and $x_{\mathcal{J}_{k-1}}$ is fixed at the iteration $k-1$ and k. We have $\tilde{W}_{k-1}\phi(x_{\tilde{W}}) = \tilde{W}_{k-1}z_{\tilde{W}} = P_{k-1}\nabla f(\tilde{W}_{k-1}x_{\tilde{W}}, x_{\mathcal{J}_{k-1}})_{\mathcal{I}_{k-1}}$. We assume that ∇f is a continuously differentiable function, locally invertible in the intersection of Ω_{k-1} with a neighborhood of $x^{(k-1)}$ including $x^{(k)}$; we define the inverse function ϕ^{-1} as $\phi^{-1}(z_{\tilde{W}}) = x_{\tilde{W}} \Leftrightarrow \phi(x_{\tilde{W}}) = z_{\tilde{W}}$, or equivalently

$$\begin{aligned} \phi^{-1}(z_{\tilde{W}}) = x_{\tilde{W}} \quad &\Leftrightarrow \quad \tilde{W}_{k-1}^T P_{k-1} \nabla f(\tilde{W}_{k-1}x_{\tilde{W}}, x_{\mathcal{J}_{k-1}})_{\mathcal{I}_{k-1}} = \\ &= \quad \tilde{W}_{k-1}^T \nabla f(\tilde{W}_{k-1}x_{\tilde{W}}, x_{\mathcal{J}_{k-1}})_{\mathcal{I}_{k-1}} = z_{\tilde{W}}. \end{aligned}$$

The Jacobian matrix of ϕ^{-1} at $z_{\tilde{W}}$ is $(\tilde{W}_{k-1}^T \nabla^2 f(\tilde{W}_{k-1}x_{\tilde{W}}, x_{\mathcal{J}_{k-1}})_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}} \tilde{W}_{k-1})^{-1}$, equal to the inverse of $\tilde{U}_{k-1}^T \nabla^2 f(x_{\mathcal{I}_{k-1}}, x_{\mathcal{J}_{k-1}})\tilde{U}_{k-1}$.
Setting $\phi^{-1}(z^{(k-1)}_{\tilde{W}}) = x^{(k-1)}_{\tilde{W}}$, with $x^{(k-1)}_{\mathcal{I}_{k-1}} = \tilde{W}_{k-1}x^{(k-1)}_{\tilde{W}}$ and $\phi^{-1}(z^{(k)}_{\tilde{W}}) = x^{(k)}_{\tilde{W}}$, with $x^{(k)}_{\mathcal{I}_{k-1}} = \tilde{W}_{k-1}x^{(k)}_{\tilde{W}}$, we can write

$$x^{(k)}_{\tilde{W}} - x^{(k-1)}_{\tilde{W}} = \int_0^1 (\tilde{U}_{k-1}^T \nabla^2 f(x^{(k-1)} + \varrho s^{(k-1)})\tilde{U}_{k-1})^{-1} (z^{(k)}_{\tilde{W}} - z^{(k-1)}_{\tilde{W}}) d\varrho.$$

By multiplying both the members of the previous equality for ${y^{(k-1)}_{\mathcal{I}_{k-1}}}^T \tilde{W}_{k-1}$, we have

$$\begin{aligned} {y^{(k-1)}_{\mathcal{I}_{k-1}}}^T s^{(k-1)}_{\mathcal{I}_{k-1}} &= {y^{(k-1)}_{\mathcal{I}_{k-1}}}^T \tilde{W}_{k-1} \left(x^{(k)}_{\tilde{W}} - x^{(k-1)}_{\tilde{W}}\right) \\ &= {y^{(k-1)}_{\mathcal{I}_{k-1}}}^T \tilde{W}_{k-1} \int_0^1 (\tilde{U}_{k-1}^T \nabla^2 f(x^{(k-1)} + \varrho s^{(k-1)})\tilde{U}_{k-1})^{-1} \tilde{W}_{k-1}^T y^{(k-1)}_{\mathcal{I}_{k-1}} d\varrho. \end{aligned}$$

Since ${t^{(k-1)}_{\mathcal{I}_{k-1}}}^T t^{(k-1)}_{\mathcal{I}_{k-1}} = {y^{(k-1)}_{\mathcal{I}_{k-1}}}^T \tilde{W}_{k-1}\tilde{W}_{k-1}^T y^{(k-1)}_{\mathcal{I}_{k-1}}$, we can conclude that $\alpha_k^{\text{EQ-BB2}}$ can be interpreted as a Rayleigh quotient relative to the average inverse of the matrix $\tilde{U}_{k-1}^T \nabla^2 f(x^{(k-1)} + \varrho s^{(k-1)})\tilde{U}_{k-1}$.

Remark 3.1. Inspired by the idea at the basis of the classical BB rules, we observe that the steplength selections (22) and (23) can be interpreted also as solutions of the following modified secant conditions:

$$\alpha_k^{\text{BB1}} = \arg\min_{\alpha\in\mathbb{R}} \|\alpha^{-1}s^{(k-1)}_{\mathcal{I}_{k-1}} - t^{(k-1)}_{\mathcal{I}_{k-1}}\|, \quad \alpha_k^{\text{EQ-BB2}} = \arg\min_{\alpha\in\mathbb{R}} \|s^{(k-1)}_{\mathcal{I}_{k-1}} - \alpha t^{(k-1)}_{\mathcal{I}_{k-1}}\|. \tag{45}$$

4. Variable metric gradient projection method

In order to improve the convergence rate of algorithm (2), a very popular technique [10, 12, 14, 36] consists in exploiting a variable metric instead of the standard fixed Euclidean one, by introducing the variable metric gradient projection method

$$d^{(k)} = \Pi_\Omega^{D_k}\left(x^{(k)} - \alpha_k D_k^{-1}\nabla f(x^{(k)})\right) - x^{(k)}, \qquad x^{(k+1)} = x^{(k)} + \nu_k d^{(k)}, \tag{46}$$

where $\{D_k\}_{k\in\mathbb{N}}$ is a sequence of symmetric and positive definite matrices with the eigenvalues belonging to $\left[\frac{1}{\mu},\mu\right]$, $\mu\geq 1$ and $\Pi_\Omega^{D_k}$ represents the projection operator onto Ω with respect to the norm induced by D_k:

$$\Pi_\Omega^{D_k}(z)=\arg\min_{x\in\Omega}\frac{1}{2}\|x-z\|_{D_k}^2\equiv\frac{1}{2}\left(x-z\right)^T D_k\left(x-z\right).$$

The selection of the sequence $\{D_k\}_{k\in\mathbb{N}}$ usually must aim at two main goals: improving the convergence rate and adding some local information about the problem without introducing significant computational costs. Typically, the definition of proper scaling matrices is strictly related to the problem to handle and, for this reason, their setting criteria will be discussed case by case in the section devoted to the numerical experiments. However, we point out that, to keep the computational cost unchanged, the scaling matrices must have a simple structure and therefore, hereafter, we will consider diagonal matrices. From a theoretical point of view, the convergence of the variable metric gradient projection method (46) is still ensured for any value of the steplength α_k belonging to a compact subset of $\mathbb{R}_+$ [9, Theorem 2.1]. This allows to properly modify the steplength selection rule in order to consider the presence of the variable metric, without being subject to restrictive conditions. A natural way to achieve this goal consists in asking α_k to satisfy generalized secant conditions written in terms of the norm induced by the matrix $(D_k)_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}}$,

$$\begin{aligned}\alpha_k^{\text{P-BB1}}&=\arg\min_{\alpha\in\mathbb{R}}\|\alpha^{-1}s_{\mathcal{I}_{k-1}}^{(k-1)}-(D_k^{-1})_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}}t_{\mathcal{I}_{k-1}}^{(k-1)}\|_{(D_k)_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}}}\\ \alpha_k^{\text{P-EQ-BB2}}&=\arg\min_{\alpha\in\mathbb{R}}\|s_{\mathcal{I}_{k-1}}^{(k-1)}-\alpha(D_k^{-1})_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}}t_{\mathcal{I}_{k-1}}^{(k-1)}\|_{(D_k)_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}}},\end{aligned}\tag{47}$$

that provides the following updating rules

$$\alpha_k^{\text{P-BB1}}=\frac{{s_{\mathcal{I}_{k-1}}^{(k-1)}}^T(D_k)_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}}s_{\mathcal{I}_{k-1}}^{(k-1)}}{{s_{\mathcal{I}_{k-1}}^{(k-1)}}^T t_{\mathcal{I}_{k-1}}^{(k-1)}}=\frac{{s^{(k-1)}}^T D_k s^{(k-1)}}{{s^{(k-1)}}^T y^{(k-1)}},\tag{48}$$

$$\alpha_k^{\text{P-EQ-BB2}}=\frac{{s_{\mathcal{I}_{k-1}}^{(k-1)}}^T t_{\mathcal{I}_{k-1}}^{(k-1)}}{{t_{\mathcal{I}_{k-1}}^{(k-1)}}^T(D_k^{-1})_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}}t_{\mathcal{I}_{k-1}}^{(k-1)}}.\tag{49}$$

The rule (49) is the modified version of the following strategy

$$\alpha_k^{\text{P-BB2}}=\frac{{s^{(k-1)}}^T y^{(k-1)}}{{y^{(k-1)}}^T D_k^{-1}y^{(k-1)}},\tag{50}$$

which takes into account the presence of the scaling matrix but does not consider the inactive constraints of the feasible region at each iteration. It is interesting to observe that, when ${s^{(k-1)}}^T y^{(k-1)}>0$, from the inequality

$$\left({s_{\mathcal{I}_{k-1}}^{(k-1)}}^T t_{\mathcal{I}_{k-1}}^{(k-1)}\right)^2\leq\left({s_{\mathcal{I}_{k-1}}^{(k-1)}}^T(D_k)_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}}s_{\mathcal{I}_{k-1}}^{(k-1)}\right)\left({t_{\mathcal{I}_{k-1}}^{(k-1)}}^T(D_k^{-1})_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}}t_{\mathcal{I}_{k-1}}^{(k-1)}\right),$$

we easily obtain that $\alpha_k^{\text{P-EQ-BB2}}\leq\alpha_k^{\text{P-BB1}}$. This suggests that the rule (48) and (49) are suitable to be tested within a strategy generalizing (43) to the variable metric case:

$$\alpha_k^{\text{P-EQ-VABB}_{\min}}=\begin{cases}\min\{\alpha_j^{\text{P-EQ-BB2}}\ :\ j=\max\{1,k-m_a\},\dots,k\} & \text{if } \frac{\alpha_k^{\text{P-EQ-BB2}}}{\alpha_k^{\text{P-BB1}}}<\tau_k,\\ \alpha_k^{\text{P-BB1}} & \text{otherwise,}\end{cases}\tag{51}$$

where m_a and τ_k are defined as in (6).

Remark 4.1. In the special case of quadratic objective functions and fixed scaling matrices, the steps of the scheme (46) are clearly related to the steps of a standard gradient method applied to a preconditioned quadratic problem. Then it would be interesting to investigate possible interpretations of the above variable metric gradient projection method as a preconditioned scheme and to evaluate its behaviour in comparison with other preconditioned approaches available in literature (e.g., [29, Section 4] and [39]). However, these topics deserve a deep analysis and are beyond the aims of this work.

5. Numerical experiments

This section is devoted to evaluate the effects of the BB-like rules previously described on the gradient projection method (2). The variable metric variant of GP, equipped with the steplength rules proposed in Section 4, is denoted by SGP. We consider several optimization problems, with both quadratic and non-quadratic objective functions.

Before presenting the results, we recall some common features shared by all the numerical experiments performed: for the GP and the SGP methods, we made the following choices.

- The projection onto the set Ω is formulated as a root-finding problem and effectively computed by the secant-like algorithm developed in [18].
- In order to guarantee the boundedness of the steplengths, we have always requested that α_k belongs to the interval $[\alpha_{\min}, \alpha_{\max}] = [10^{-10}, 10^6]$.
- The parameter ν_k has been selected by means of the non-monotone version of the Armijo linesearch proposed in [30]: for given scalars $0 < \beta, \delta < 1$, the parameter ν_k is set equal to β^{i_k} and i_k is the first non-negative integer i for which

$$f_{\max} - f(x^{(k)} + \beta^i d^{(k)}) \geq -\delta \beta^i \nabla f(x^{(k)})^T d^{(k)}\,, \tag{52}$$

  where $f_{\max}$ is the maximum value of the objective function in the last M iterations. For all the experiments we set $\beta = 0.4$, $\delta = 10^{-4}$ and $M = 10$.

  We recall that every accumulation point of the sequence $\{x^{(k)}\}_{k\in\mathbb{N}}$ generated by either algorithm (2) or its variable metric variant (46), equipped with the non-monotone linesearch (52), is a stationary point of (1), as proved in [10, Theorem 2.1].

All the numerical experiments are performed in the Matlab environment.

5.1. Random SLB problems

In this section, we consider randomly generated test problems. For the quadratic case, we analyze the behaviour of the BB strategies and their modified versions on some toy problems, in order to show graphically the validity of inequalities (24) and (25). Furthermore, for the non-quadratic case, we perform a study on a set of 162 well-known test problems of both small and large dimensions.

5.1.1. Quadratic case

The aim of this section is both to verify the efficiency of the modified BB2 steplength rule employed alone or in an alternating scheme and to analyze the distribution of the steplengths with respect to the eigenvalues of the sequence of the restricted Hessian matrices obtained during the iterative process. In order to reach these goals we randomly generated quadratic SLB problems where the solution, the vector v, the number of active constraints at the solution and the distribution of the eigenvalues of the Hessian matrix of the objective function are prefixed. Table 1 summarizes the main features of the three test problems we considered for the investigation of this section.

We evaluate the behaviour of the gradient projection method (2) equipped with different steplength selection rules: BB1, BB2, EQ-BB2, $\text{ABB}_{\min}$, EQ-$\text{ABB}_{\min}$ and EQ-$\text{VABB}_{\min}$. We recall that $\text{ABB}_{\min}$ and EQ-$\text{ABB}_{\min}$ stand for the particular cases of (6) and (43) where $\tau_k = \tau,\ \forall k$; in our experiments we

Table 1: Features of the randomly generated quadratic SLB problems.

	n	$\lambda_{\min}(A)$	$\lambda_{\max}(A)$	Distribution of the eigenvalues of A	$\lambda_{\min}(A^*)$	$\lambda_{\max}(A^*)$
QP1	1000	19	9923	Marchenko-Pastur [38, 46]	41.09	9862.17
QP2	1000	1	1000	Log-spaced	5.29	679.87
QP3	500	0.05	1455.57	Log-spaced	10.03	99.86

set $\tau = 0.7$. On the other hand, for the EQ-VABB$_{\min}$ scheme, we fixed $\tau_1 = 0.7$, $m_a = 2$ and $\zeta = 1.3$. The following stopping criterion was used to stop the GP methods:

$$\|x^{(k)} - x^{(k-1)}\|_\infty \leq 10^{-7}\,. \tag{53}$$

Figure 1 shows the behaviour of the inverses of the steplengths generated by means of BB1, BB2 and EQ-BB2 with respect to the eigenvalues of the matrices $\tilde{U}_{k-1}^T A \tilde{U}_{k-1}$. Particularly, in the first three rows of Figure 1, at the k-th iteration, we plotted, by black dots, 20 eigenvalues of the restricted Hessian matrix $\tilde{U}_{k-1}^T A \tilde{U}_{k-1}$ with linearly spaced indices (always including the maximum and minimum eigenvalues) and, by a red cross, the inverse of the steplength α_k. Moreover, the blue lines correspond to the maximum and the minimum eigenvalues of the whole Hessian matrix A and the blue circles denote the 20 eigenvalues of the restricted Hessian matrix at the prefixed solution x^*, with the same linearly spaced indices considered for the plot at the single iteration. These plots confirm that the inverses of the steplengths generated by the BB1 and the EQ-BB2 rules satisfy inequalities (24)-(25), while the inverses of the steplengths obtained by applying the non-modified BB2 scheme can fall outside the spectrum of the restricted Hessian matrices. The fourth row of Figure 1 shows the decrease of the following relative distance

$$\frac{|f(x^{(k)}) - f^*|}{|f^*|} \tag{54}$$

between the objective function values obtained by considering the different BB-like rules in the GP method and f^*, namely the objective function value at the solution. The decrease of the objective function towards the minimum is considerably accelerated by employing the EQ-BB2 strategy instead of the BB2 one. The alternating steplength selection rules also take advantage of the modified version of the BB2 scheme, which accounts for the nature of the problem feasible set.

5.1.2. Non-quadratic case

In this section we analyze the practical efficiency of the considered methods on some non-quadratic SLB test problems.

Based on the technique proposed in [24], the test problems were generated in two possible ways, as follows. Starting from an unconstrained minimization problem with a twice continuously differentiable objective function $\phi(x)$,

$$\min_{x \in \mathbb{R}^n} \phi(x), \tag{55}$$

for which a local minimum point x^* is known, we generated a constrained problem having one of the following formulations:

$$\begin{aligned} &\min_{x \in \mathbb{R}^n} f(x) = \phi(x) + v^T(x - x^*) + \sum_{i \in L} h_i(x_i) - \sum_{i \in U} h_i(x_i) \\ &\text{subject to } \ \ell \leq x \leq u, \quad v^T x = e, \end{aligned} \tag{56}$$

or

$$\begin{aligned} &\min_{x \in \mathbb{R}^n} f(x) = \phi(x) v^T x + \sum_{i \in L} h_i(x_i) - \sum_{i \in U} h_i(x_i) \\ &\text{subject to } \ \ell \leq x \leq u, \quad v^T x = e, \quad e > 0, \end{aligned} \tag{57}$$

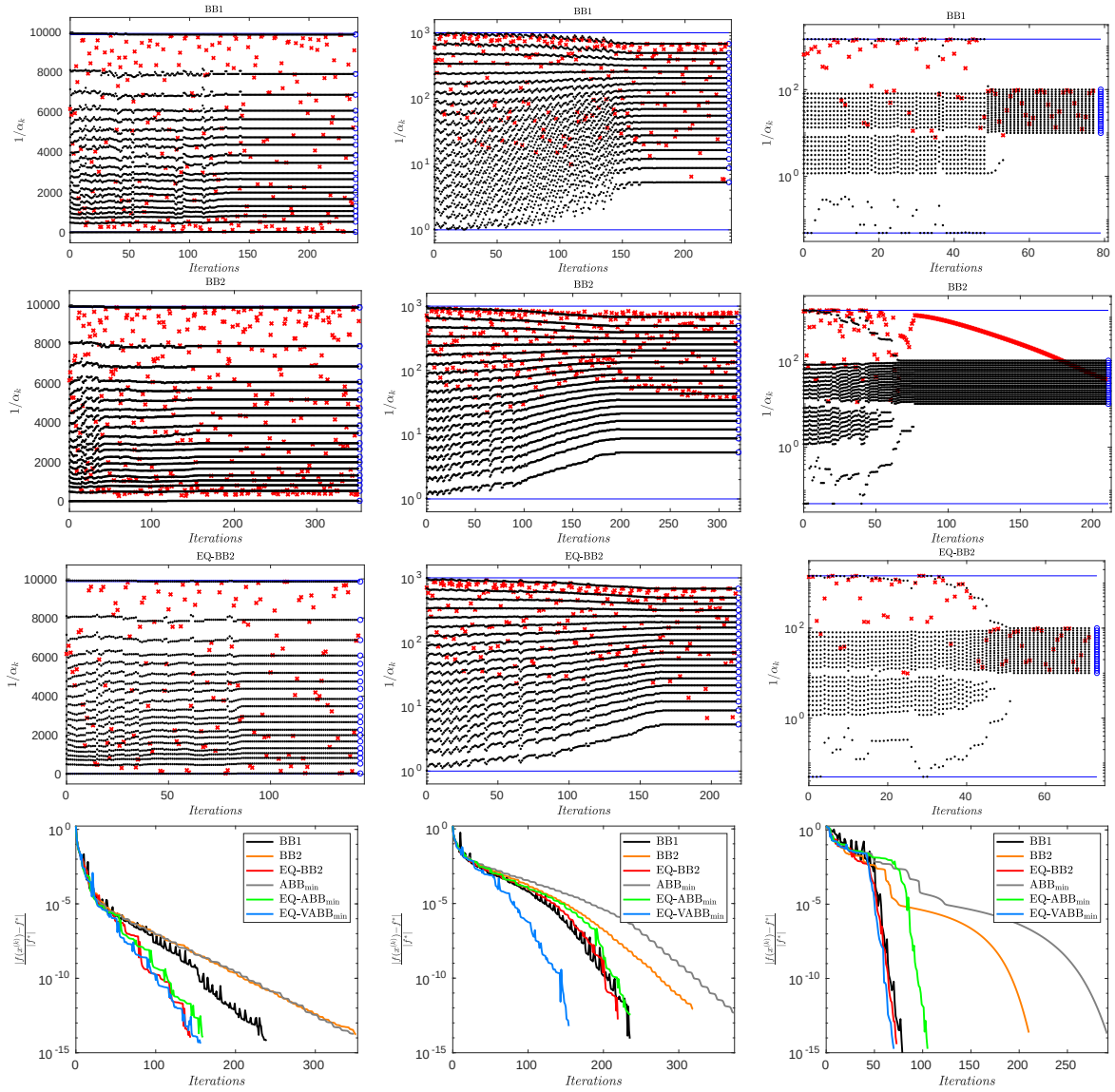

Figure 1: Behaviour of gradient projection method equipped with different steplength rules on QP1 (left column), QP2 (middle column), QP3 (right column). Distribution of α_k with respect to the iterations for the BB1 (first row), BB2 (second row), EQ-BB2 (third row) rules; error on $f(x^{(k)})$ for the different rules (fourth row).

where $v \in \mathbb{R}^n$, $e \in \mathbb{R}$, $L = \{i \,|\, x_i^* = \ell_i\}$, $U = \{i \,|\, x_i^* = u_i\}$ and $h_i \colon \mathbb{R} \to \mathbb{R}$, $i \in L \cup U$, are twice continuously differentiable non-decreasing functions. Note that the constrained problems defined by (56) and (57) have the same solution x^* of the unconstrained problem (55); to this end, the scalar e in the second formulation must be positive. For our tests we selected some well-known non-quadratic functions $\phi(x)$, described below.

(i) Trigonometric function [27]:

$$\phi(x) = \|b - (A\tilde{d}(x) + B\tilde{q}(x))\|^2,$$

where $\tilde{d}(x) = (\sin(x_1), ..., \sin(x_n))^T$, $\tilde{q}(x) = (\cos(x_1), ..., \cos(x_n))^T$, and A and B are square matrices of order $n = 500$ with entries generated as random integers in $(-100, 100)$. Given a vector $x^* \in \mathbb{R}^n$ with entries randomly generated from a uniform distribution in $(-\pi, \pi)$, the vector b is defined so that $\phi(x^*) = 0$.

(ii) Chained Rosenbrock function [44]:

$$\phi(x) = \sum_{i=2}^{n} (4\varphi_i(x_{i-1} - x_i^2)^2 - (1 - x_i)^2),$$

where $n = 500$, the values $\varphi_i,\ i = 1, \ldots, 50$, are defined as in [44, Table 1] and $\varphi_{i+50j} = \varphi_i$, $i = 1, ..., 50,\ j = 1, \ldots, 9$. In this case, a solution of the problem (55) is $x^* = (1, 1, ..., 1)^T$.

(iii) Laplace2 function [26]:

$$\phi(x) = \frac{1}{2} x^T A x - b^T x + \frac{1}{4} h^2 \sum_i x_i^4,$$

where A is a square matrix of order $n = N^3$, $N = 100$, arising from the discretization of a 3D Laplacian on the unit box by a standard seven-point finite difference formula, $h = \frac{1}{N+1}$ and b is chosen so that

$$x_i^* \equiv x(kh, rh, sh) = h^3 krs(kh - 1)(rh - 1)(sh - 1)e^{-\frac{1}{2}}((kh - d_1)^2 + (rh - d_2)^2 + (sh - d_3)^2),$$

where the index i is associated with the mesh point (kh, rh, sh), $k, r, s = 1, \ldots, N$ in a lexicographic ordering. Two different settings for the parameters d, d_1, d_2 and d_3 are considered:

a) $d = 20$, $d_1 = d_2 = d_3 = 0.5$,

b) $d = 50$, $d_1 = 0.4$, $d_2 = 0.7$, $d_3 = 0.5$.

We used both the formulations (56) and (57) to build the corresponding constrained versions of test problems (i) and (ii), whereas for test problems (iii) we used the latter form only. In our tests, v was randomly generated from a uniform distribution in $(0, 1)$ and for each of the functions $\phi(x)$ the following choices for the functions $h_i(x)$ were made, as suggested in [24]:

(1) $\beta_i\,(x_i - x_i^*)$,

(2) $\alpha_i\,(x_i - x_i^*)^3 + \beta_i\,(x_i - x_i^*)$,

(3) $\alpha_i\,(x_i - x_i^*)^{7/3} + \beta_i\,(x_i - x_i^*)$,

where α_i are random numbers in $(0.001, 0.011)$ and $\beta_i = 10^{-\eta_i ndeg}$, with η_i random numbers in $(0, 1)$ and $ndeg = 1, 4, 10$. In order to retain first-order optimality conditions, the Lagrangian multiplier of the single linear equality constraint must be equal to 1 up to sign (for the case (56)) or to $\phi(x^*)$ up to sign (for the case (57)), while the Lagrangian multipliers associated to the active constraints are easily assigned equal to the values β_i, and, therefore, the parameter $ndeg$ allows to control the degeneracy of the problem at x^*. The vectors ℓ and u were defined in order to have the number of active constraints at the solution equal to a prefixed value na; in particular, we set $na \approx 0.1 \cdot n, 0.5 \cdot n, 0.9 \cdot n$ and the same number of lower and upper active constraints at x^*. The resulting dataset is composed of 162 non-quadratic SLB test problems. We evaluated the performance obtained by running the GP method equipped with the steplengths rules: BB1, BB2, ABB$_{min}$, EQ-BB2, EQ-ABB$_{min}$, EQ-VABB$_{min}$. The considered schemes shared the following parameter setting: $\alpha_0 = 1$, $\tau = 0.7$ and $m_a = 2$ for defining $\alpha_k^{\text{ABB}_{\min}}$ and $\alpha_k^{\text{EQ-ABB}_{\min}}$ and $\tau_1 = 0.7$, $m_a = 2$ and $\zeta = 1.3$ for $\alpha_k^{\text{EQ-VABB}_{\min}}$. As for the starting vector, we considered

- $x^{(0)} = \Pi_\Omega\,(x^* + 0.3\,r)$, where $r \in \mathbb{R}^n$ has random entries from a uniform distribution in $[-\pi, \pi]$, when $\phi(x)$ is the trigonometric function previously defined in (i);
- $x^{(0)} = \Pi_\Omega(x^* + 0.8\,r)$, where $r \in \mathbb{R}^n$ has random entries from a uniform distribution in $[-1, 1]$, when $\phi(x)$ is the Chained Rosenbrock function previously defined in (ii);
- $x^{(0)} = \Pi_\Omega\left(\frac{\ell+u}{2}\right)$, when $\phi(x)$ is the Laplace2 function previously defined in (iii).

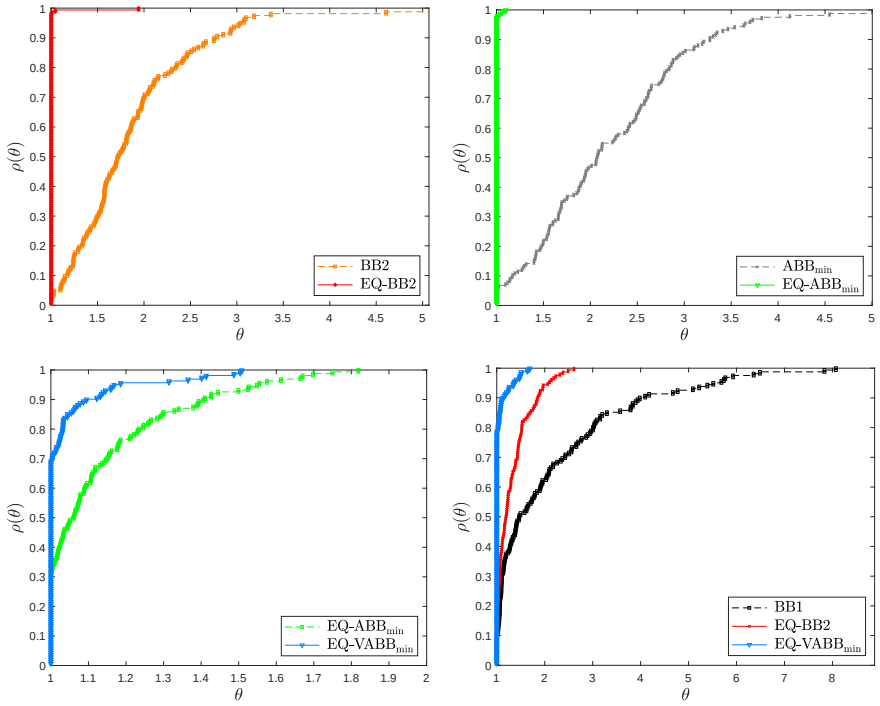

Figure 2: Runtime performance profiles of the GP method equipped with different steplength rules on a set of 162 non-quadratic SLB test problems.

In these tests, we adopted the stopping criterion (53). The results obtained were compared using the performance profiles proposed in [23]. We assumed as performance measure of interest the execution time required by each scheme to satisfy the stopping criterion, declaring a failure when it was not fulfilled within the maximum number of 4000 iterations. Some performance profiles of different GP versions are reported in Figure 2. In particular, each profile $\rho(\theta)$ gives the fraction of the problems that a solver is able to solve within a factor θ of the best time of all the solvers; thus, $\rho(1)$ represents the fraction of problems for which the considered solver is the winner, while the performance profile corresponding to the largest value of θ gives the fraction of problems for which the considered solver is successful.

The performance profiles generated for the random non-quadratic SLB problems confirm the results obtained in the quadratic framework. By capturing the information about the active set at each iteration, the EQ-BB2 steplength strategy allows the GP method to improve its behaviour in terms of computational time with respect to the case in which the original BB2 rule is used. This improvement still holds if we employ the alternating strategies equipped with the new EQ-BB2 rule in place of the BB2 one.

5.2. SLB problems from real-life applications

In this section we evaluate the performance of the gradient projection method with both standard and modified BB strategies in solving SLB problems arising from three different real-life applications.

5.2.1. Support Vector Machines

The learning methodology called support vector machines (SVMs) implies to solve a quadratic SLB problem. In order to delineate the features of such a problem, we briefly recall the SVMs framework (we refer the reader to [11, 15] for a detailed discussion). If $D = \{(z_i, y_i), \quad i = 1, \ldots, n, \quad z_i \in \mathbb{R}^m, \quad y_i \in \{-1, 1\}\}$ is a training set of labeled examples, the SVM algorithm performs classification of new examples

$z \in \mathbb{R}^m$ by using a decision function $F : \mathbb{R}^m \longrightarrow \{-1, 1\}$ of the form

$$F(z) = \text{sign}\left(\sum_{i=1}^{n} x_i^* y_i K(z, z_i) + b^*\right), \tag{58}$$

where $K \colon \mathbb{R}^m \times \mathbb{R}^m \longrightarrow \mathbb{R}$ denotes a kernel function and $x^* = (x_1^*, \ldots, x_n^?)^T$ is the solution of

$$\begin{aligned} &\min f(x) = \frac{1}{2} x^T A x - \sum_{i=1}^{n} x_i \\ &\text{subject to } \; 0 \le x \le C, \quad \sum_{i=1}^{n} y_i x_i = 0\,. \end{aligned} \tag{59}$$

Once the vector x^* is computed, the quantity b^* in (58) is easily derived. The Hessian matrix A of (59) has entries $A_{ij} = y_i y_j K(z_i, z_j), \quad i, j = 1, 2, \ldots, n$ and $C \in \mathbb{R}^n$ is a vector with all entries equal to a positive parameter. For our test problems we consider a Gaussian kernel, namely $K(z_i, z_j) = e^{-\frac{\|z_i - z_j\|_2^2}{2\sigma^2}}$, with $\sigma \in \mathbb{R}$. In order to appreciate the validity of the considerations made in Section 2, we compare the behaviour of the gradient projection method (2) by varying the steplength among α_k^{BB1}, α_k^{BB2}, $\alpha_k^{\text{VABB}_{\min}}$, $\alpha_k^{\text{EQ-BB2}}$, $\alpha_k^{\text{EQ-VABB}_{\min}}$ in solving problem (59) for four different datasets with the following features:

MNIST1000 $n = 1000$, $C = 10$, $\sigma = 1800$, $\text{rank}(A) = 1000$;

MNIST2000 $n = 2000$, $C = 10$, $\sigma = 1800$, $\text{rank}(A) = 2000$;

ADU $n = 1000$, $C = 1$, $\sigma = \sqrt{10}$, $\text{rank}(A) = 985$;

WEB $n = 1000$, $C = 5$, $\sigma = \sqrt{10}$, $\text{rank}(A) = 736$.

We generated these datasets starting from a learning problems repository, called LIBSVM, available at https://www.csie.ntu.edu.tw/∼cjlin/libsvmtools/datasets/. We selected some datasets for the binary classification and we set the dimension by ourselves. For the alternating steplength selection rules we adopted $\tau_1 = 0.7$, $m_a = 2$ and $\zeta = 1.3$. Finally, the GP method (2) has been stopped when either the relative distance between two successive iterations was lower than 10^{-8} or 1000 iterations have been performed. The initial point for all the considered schemes is the null vector.
Given the smallest value f^* of the objective function among the ones obtained by the different methods at the end of the iterative process, Table 2 shows the number of iterations and the computational time needed by the considered schemes to reduce the relative difference (54) below a certain tolerance *tol*. If a method does not succeed in realizing this goal in the prefixed maximum number of iterations (1000), the corresponding entry of the table reports the minus sign. In Figure 3, we can appreciate the decrease of the relative error (54), with respect to the computational time, for the four datasets. The results reported in Table 2 and Figure 3 confirm the effectiveness of the modified BB2 selection rule with respect to the original one, also within the alternating scheme. The benefits gained by employing EQ-BB2 instead of BB2 are clear in terms of both number of iterations and computational time.

5.2.2. Reconstruction of fiber orientation distribution in diffusion MRI

The aim of this section is to consider the problem of intra-voxel reconstruction of the fiber orientation distribution function (FOD) in each voxel of the white matter of the brain from diffusion MRI data. In [34] the authors clarify that the diffusion signal can be represented as the convolution of a response function with the FOD function and, as a consequence, the estimation of the intra-voxel structure can be shaped through a linear model of the form

$$b = \Phi x + \eta, \tag{60}$$

Table 2: Number of iterations required by each algorithm to reduce the difference (54) below given tolerances for the four SVM test problems. The corresponding computational time (averaged over 20 runs) is also reported.

	$tol = 10^{-2}$		$tol = 10^{-4}$		$tol = 10^{-6}$		$tol = 10^{-8}$	
	It.	Time	It.	Time	It.	Time	It.	Time
				MNIST1000				
BB1	47	0.016	91	0.046	140	0.061	177	0.072
BB2	88	0.025	223	0.060	297	0.079	389	0.102
EQ-BB2	41	0.014	88	0.027	114	0.035	146	0.056
VABB_{min}	48	0.017	92	0.029	145	0.043	175	0.064
$\text{EQ-VABB}_{\text{min}}$	36	0.013	81	0.026	121	0.038	142	0.044
				MNIST2000				
BB1	77	0.136	172	0.299	285	0.500	418	0.732
BB2	171	0.314	458	0.789	917	1.558	-	-
EQ-BB2	78	0.138	153	0.272	219	0.381	268	0.460
VABB_{min}	71	0.114	162	0.255	250	0.398	337	0.538
$\text{EQ-VABB}_{\text{min}}$	57	0.096	130	0.219	186	0.313	233	0.392
				ADU				
BB1	23	0.009	45	0.017	79	0.028	113	0.040
BB2	30	0.010	73	0.022	163	0.047	322	0.098
EQ-BB2	22	0.008	47	0.016	72	0.024	104	0.034
VABB_{min}	29	0.010	67	0.023	109	0.035	164	0.051
$\text{EQ-VABB}_{\text{min}}$	23	0.009	60	0.022	85	0.031	111	0.040
				WEB				
BB1	84	0.027	254	0.082	519	0.163	755	0.234
BB2	315	0.079	1000	0.304	-	-	-	-
EQ-BB2	59	0.020	193	0.062	338	0.107	468	0.155
VABB_{min}	103	0.030	260	0.075	488	0.141	767	0.223
$\text{EQ-VABB}_{\text{min}}$	60	0.020	162	0.053	343	0.112	407	0.133

where $x \in \mathbb{R}^n$ represents the FOD function, $b \in \mathbb{R}^m$ is the vector of measurements, Φ is the linear measurement operator, and η is the acquisition noise. Since problem (60) is ill-posed, it has been proved [1] to be convenient finding a meaningful solution by means of a reweighted ℓ_1-minimization process which involves at each step the solution of a convex problem of the form

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} f(x) \equiv \|\Phi x - b\|_2^2 \\ &\text{subject to} \quad x \geq 0, \quad \|Wx\|_1 = K, \end{aligned} \tag{61}$$

where $W \in \mathbb{R}^{n\times n}$ is a diagonal matrix with positive entries and K is the estimated maximum number of fibres to be detected in the brain volume. The weighted ℓ_1-norm constraint induces sparsity on the solution and the weighting matrix W forces some anatomical properties of the fiber bundles in neighboring voxels. A complete overview about the properties of W can be found in [1].

It has been shown in [13] and [8] that the presence of a variable metric in first order methods can significantly improve the performance in solving problem (61) with respect to their standard non-scaled versions. For this reason we only report the results obtained by comparing the variable metric gradient projection method SGP with different choices for the steplength parameter. Particularly, we consider the P-BB1 and P-BB2 defined in (48) and (50) respectively, the modified version of P-BB2 fixed in (49), called P-EQ-BB2, and the alternating strategies $\text{P-EQ-VABB}_{\text{min}}$ set in (51) and $\text{P-VABB}_{\text{min}}$, which can be obtained by (51) by considering the P-BB2 rule instead of the P-EQ-BB2. For $\text{P-EQ-VABB}_{\text{min}}$ and $\text{P-VABB}_{\text{min}}$ the parameters m_a, τ_1 and ζ have been chosen equal to 2, 0.5 and 3, respectively. As for the variable metric, the sequence $\{D_k\}_{k\in\mathbb{N}}$ has been selected by mimicking the split gradient-based scaling proposed in [3] for quadratic problems: the scaling matrix has the following form

$$D_k = \text{diag}\left(\max\left(\frac{1}{\mu_k}, \min\left(\mu_k, \frac{x^{(k)}}{\Phi^T\Phi x^{(k)}}\right)\right)\right)^{-1}, \tag{62}$$

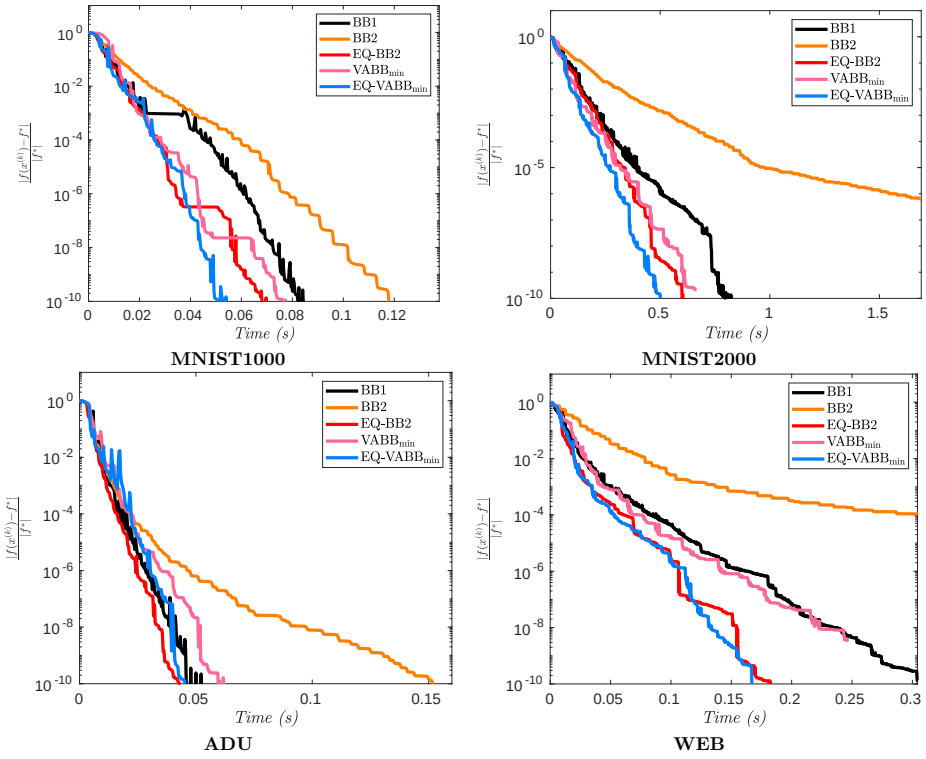

Figure 3: Plots of the relative difference (54) with respect to the computational time for the four SVM test problems

where $\mu_k = \sqrt{1 + \frac{10^{11}}{(k+1)^2}}$. Thanks to the parameter μ_k we force the sequence $\{D_k\}_{k\in\mathbb{N}}$ to asymptotically approach the identity matrix [14, Lemma 2.3]. This condition ensures the convergence of the sequence of the iterates generated by the SGP scheme to a solution of the minimization problem, as proved in [9, Theorem 3.1].

For the numerical comparison we employed the Phantom dataset, which is available at https://github.com/basp-group/co-dmri and is described in [1]. In particular, for this test problem the parameters m, n, K have the following values: $m = 19200$, $n = 257280$, and $K = 3840$. Table 3 summarizes the number of iterations and the computational time needed by the considered scheme to guarantee that the distance (54) is below certain thresholds *tol*. If this requirement is not satisfied in the prefixed maximum number of iterations (4000) the corresponding entry reports the minus sign Figure 4 shows the relative difference (54) between the objective function values provided by the different methods and the minimum computed value f^*.

By analyzing the results offered in Table 3 and Figure 4, we can reach analogous conclusions to the ones made for the previous numerical experiments. The modified version of the BB2 rule allows the SGP algorithm to largely improve its behaviour in terms of number of iterations and computational time with respect to the performance shown when it is combined with the standard BB2 strategy. Similar considerations can be done by comparing the alternating schemes: the use of EQ-BB2 instead of BB2 makes P-EQ-VABB$_{\min}$ much more effective than P-VABB$_{\min}$ in finding the solution of the optimization problem.

5.2.3. Image deblurring with Poisson noise

In order to evaluate the behaviour of the proposed steplength rules for a non-quadratic problem, we consider the reconstruction of an image b corrupted by Poisson noise by a smooth TV regularization approach. In a Bayesian framework, an approximation of the original object can be obtained by solving a constrained problem where the objective function is the sum of a discrepancy function, typically

Table 3: Number of iterations and computational time required by each algorithm to reduce the difference (54) below given tolerances for problem (61). The corresponding computational time is also reported.

	$tol = 10^{-3}$		$tol = 10^{-5}$		$tol = 10^{-8}$		$tol = 10^{-10}$	
	It.	Time	It.	Time	It.	Time	It.	Time
P-BB1	263	2.7	1025	10.6	3541	38.7	-	-
P-BB2	620	6.4	-	-	-	-	-	-
P-EQ-BB2	282	3.1	830	9.4	2552	29.9	-	-
P-VABB$_{\text{min}}$	283	2.9	3175	36.0	-	-	-	-
P-EQ-VABB$_{\text{min}}$	271	3.1	683	8.1	2169	26.5	3034	37.7

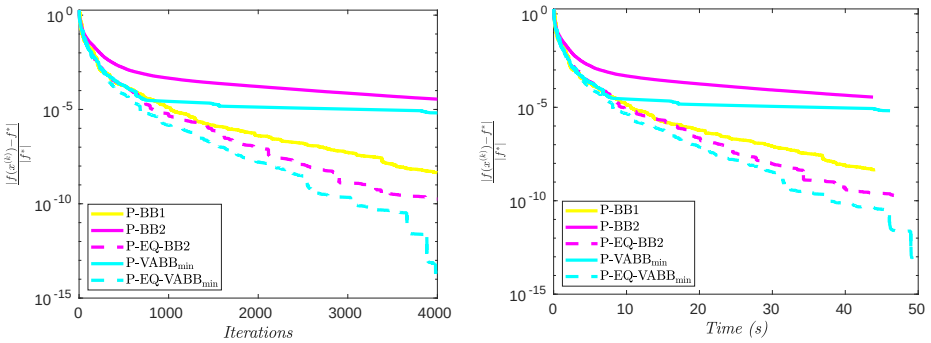

Figure 4: Test problem (61): plots of the relative difference (54) with respect to the number of iterations (first panel) and the computational time (second panel) achieved by the SGP methods .

depending on the noise type affecting the data, and a regularization term adding *a priori* information; simple constraints, expressing physical requirements, can be considered. In the case of Poisson noise, the discrepancy function measuring the distance from the data b is the generalized Kullback-Leibler (KL) divergence, having the form

$$f_0(Ax + c; b) = \sum_{i=1}^{n} b_i \log \frac{b_i}{(Ax + c)_i} + (Ax + c)_i - b_i, \tag{63}$$

where $A \in \mathbb{R}^{n \times n}$ is a linear operator modeling the distortion due to the image acquisition system and $c \in \mathbb{R}^n$ is a known positive background radiation constant. A typical assumption for the matrix A is that it has nonnegative elements and each row and column has at least one positive entry (see [5] for the details about the image deblurring problem in presence of Poisson noise). A widely used edge-preserving regularizer is the discrete smooth Total Variation functional, known also as Hyper-surface regularizer, that for an image of $n = N \times N$ pixels is defined as

$$f_1(x) = \sum_{k,\ell} \sqrt{(x_{k+1,\ell} - x_{k,\ell})^2 + (x_{k,\ell+1} - x_{k,\ell})^2 + \gamma^2} - \gamma, \tag{64}$$

where γ is a small positive constant and periodic boundary conditions are assumed. In summary, a *maximum a posteriori* estimate of the original image is a solution of the following nonlinear programming problem

$$\min_{x \in \mathbb{R}^n} f(x) \equiv f_0(Ax + c; b) + \rho f_1(x) \quad \text{subject to} \quad x \geq 0, \quad \sum_{i=1}^{n} x_i = K, \tag{65}$$

where $K = \sum_{i=1}^{n} b_i - n \cdot c$ is the flux of the image and ρ is a positive parameter balancing the role of the regularization term and the discrepancy function; the inequality constraints and the single linear equality constraint express the non-negativity of the pixels and the conservation of the image flux, respectively.

Table 4: Image deblurring test problem: number of iterations and computational time (averaged over 10 runs) required to reduce the relative error on the objective function below a prefixed tolerance. The corresponding relative reconstruction error achieved is also reported.

	$tol = 5 \cdot 10^{-2}$			$tol = 10^{-3}$			$tol = 5 \cdot 10^{-4}$		
	It.	Time	RRE	It.	Time	RRE	It.	Time	RRE
P-BB1	64	4.8	0.568	884	61	0.447	1000*	69*	0.447*
P-BB2	28	2.1	0.531	1000*	67*	0.446*	-	-	-
P-EQ-BB2	27	2.0	0.533	1000*	69*	0.446*	-	-	-
P-VABB$_{\min}$	26	1.9	0.537	427	29	0.438	1000*	68*	0.438*
P-EQ-VABB$_{\min}$	37	3.0	0.531	264	19	0.438	590	41	0.438

In particular we can consider as test problem a 512×512 object representing a micro-tubulin network inside a cell [42]. In this case, the values of the original object $\overline{x}$ are in the range $[0, 686]$, whereas those of the blurred and noisy image b are in $[0, 446]$; the background was set equal to 1 and the relative distance between the original object and the blurred noisy data in Euclidean norm is 0.756; ρ was set equal to $4 \cdot 10^{-4}$. The value of γ was set equal to $10^{-6} \cdot \max_i\{b_i\}$. A ground-truth solution x^*, i.e., an estimate of the real minimum point of the problem (65), is obtained by executing an huge number of iterations of the SGP method in [10]. Indeed, it is well known that the above problem can be efficiently addressed by the gradient projection method equipped with a variable metric (see for example [35, 10, 12]). Mimicking the split gradient-based scaling, as in the previous section, the sequence of scaling matrices $\{D_k\}_{k\in\mathbb{N}}$ can be selected as follows

$$B_k = \operatorname{diag}\left(\max\left(\frac{1}{\mu_k}, \min\left(\mu_k, \frac{x^{(k)}}{A^T\mathbf{1} + \rho V(x^{(k)})}\right)\right)\right)^{-1}, \tag{66}$$

where $\mathbf{1}$ is a vector with all entries equal to 1, $V(x^{(k)})$ is the positive part of the splitting of $\nabla f_1(x) = V(x) - U(x)$ at $x^{(k)}$ (see [5, Cap. 5]) and $\mu_k = \sqrt{1 + \frac{10^{11}}{(k+1)^2}}$.

In Table 4 and Figure 5 we report the behaviour of the scaled gradient projection method combined with the steplength rules P-BB1, P-BB2, P-EQ-BB2, P-VABB$_{\min}$ and P-EQ-VABB$_{\min}$. For the alternating rules, we have the following setting of parameters: $m_a = 2$, $\tau_1 = 0.5$ and $\zeta = 3$.
Table 4 shows the number of iterations and the time, in seconds, required by the considered methods to reduce the relative error on the objective function (54) below a fixed threshold tol, where, in this case, f^* represents the value of f at the ground-truth x^*. We report also the relative reconstruction error (RRE) $\frac{\|x^{(\mathrm{It.})} - \overline{x}\|_2}{\|\overline{x}\|_2}$ at the iteration It. If one of the approaches is not be able to reduce the relative error on the objective function under a certain tolerance within 1000 iterations, Table 4 displays the computational time spent and the RRE achieved after the 1000 iterations performed. We denote the corresponding results by a star. Figure 5 shows the relative error of the objective function and the relative minimization error $\frac{\|x^{(k)} - x^*\|_2}{\|x^*\|_2}$ with respect to the number of iterations and the computational time.

We observe that the rules P-BB2 and P-EQ-BB2 have the same behaviour, very similar to the one of P-BB1; indeed, for this problem, the variable metric (in particular the term $x^{(k)}$ in (66)) hides the effects of the rules that take into account the constraints and, at the same time, the equality constraint plays a minor role, since the assumptions on the matrix A already induce the iterates to satisfy the flux constraint. Nevertheless, when the alternating rule is adopted, the use of P-EQ-BB2 can still improve the performance of SGP, achieving in 38 s. (537 iterations) with P-EQ-VABB$_{\min}$ the value of the objective function obtained with P-VABB$_{\min}$ after 67 s. (1000 iterations).

6. Conclusions

This paper deals with the study of the spectral properties of the well-known Barzilai and Borwein steplength selection rules, often employed to accelerate classical gradient projection methods. In the

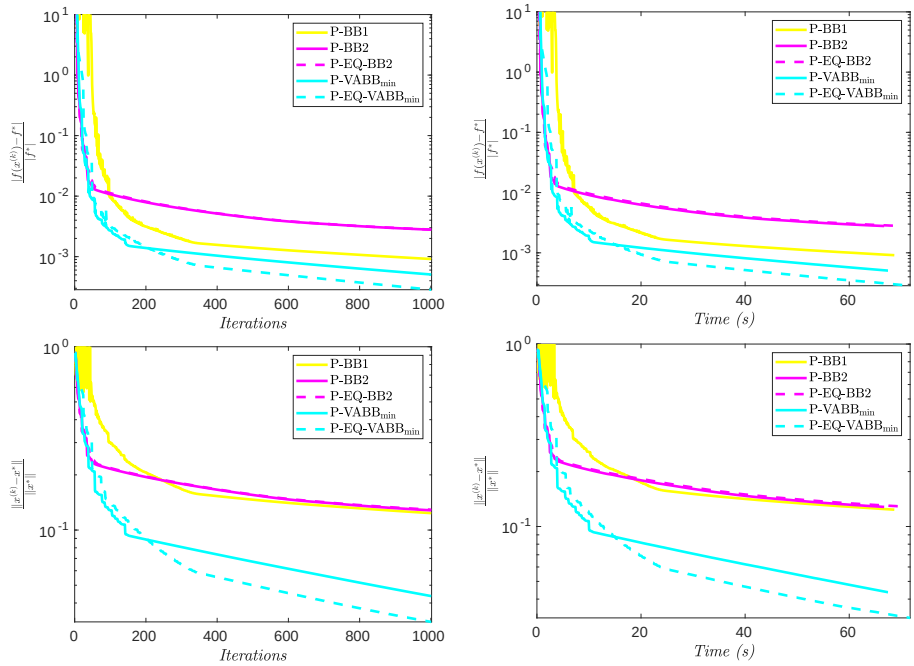

Figure 5: Image deblurring test problem: numerical results of 1000 iterations of the SGP method combined with different steplength rules; first row: relative error of the objective function with respect to the iterations (left panel) and the computational time (right panel); second row: relative minimization error $\frac{\|x^{(k)}-x^*\|}{\|x^*\|}$ with respect to the iterations (left panel) and the computational time (right panel).

literature, several works are devoted to explain how, in presence of unconstrained quadratic minimization problems, the steplength generated by the BB strategies are related to the eigenvalues of the symmetric and positive definite Hessian matrix of the objective function. This relation has been identified as the responsible for the effectiveness of the BB approaches. However the influence of a feasible set on the behaviour of the BB rules has been investigated only very recently [16] in the case of box-constrained strictly convex quadratic optimization problems. Our work represents an extension of the analysis presented in [16] for two main reasons.

- A more complicated feasible region defined by a linear equality constraint together with lower and upper bounds has been investigated. All the results developed in [16] can be seen as a particular case of the ones achieved in this paper.
- A possible interpretation of the BB steplength strategies has been also offered in the more general non-quadratic framework in terms of a proper average matrix depending on the Hessian matrix of the objective function at each iteration.

Thanks to the spectral analysis, a redefinition of one of the two BB rules has been suggested in order to take into account not only the second-order information related to the Hessian matrix of the objective function, but also the nature of the constraints. Several numerical experiments, carried out on both quadratic and non-quadratic datasets, showed the effectiveness of this modified BB strategy, also when employed in selection scheme alternating the new rule with the standard BB steplength.

References

[1] A. Auria, A. Daducci, J.-P. Thiran, and Y. Wiaux. Structured sparsity for spatially coherent fibre orientation estimation in diffusion MRI. *NeuroImage*, 115:245–255, 2015.

[2] J. Barzilai and J. M. Borwein. Two-point step size gradient methods. *IMA J. Numer. Anal.*, 8:141–148, 1988.

[3] F. Benvenuto, R. Zanella, L. Zanni, and M. Bertero. Nonnegative least-squares image deblurring: improved gradient projection approaches. *Inverse Problems*, 26:025004, 2010.

[4] M. Bertero and P. Boccacci. *Introduction to Inverse Problems in Imaging*. Institute of Physics Pub., 1998.

[5] M. Bertero, P. Boccacci, and V. Ruggiero. *Inverse Imaging with Poisson Data*. 2053-2563. IOP Publishing, 2018.

[6] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 2nd edition, 1999.

[7] E. G. Birgin, J. M. Martinez, and M. Raydan. Non-monotone spectral projected gradient methods on convex sets. *SIAM J. Optim.*, 10(4):1196–1211, 2000.

[8] S. Bonettini, F. Porta, and V. Ruggiero. A variable metric forward–backward method with extrapolation. *SIAM J. Sci. Comput.*, 38(4):A2558–A2584, 2016.

[9] S. Bonettini and M. Prato. New convergence results for the scaled gradient projection method. *Inverse Problems*, 31(9):095008, 2015.

[10] S. Bonettini, R. Zanella, and L. Zanni. A scaled gradient projection method for constrained image deblurring. *Inverse Problems*, 25(1):015002 (23pp), 2009.

[11] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.*, 2(2):121–167, 1998.

[12] E. Chouzenoux, J.-C. Pesquet, and A. Repetti. Variable metric forward–backward algorithm for minimizing the sum of a differentiable function and a convex function. *J. Optim. Theory Appl.*, 162(1):107–132, 2014.

[13] V. L. Coli, V. Ruggiero, and L. Zanni. Scaled first-order methods for a class of large-scale constrained least squares problems. In *AIP Conference Proceedings 1776*, page 040002, 2016.

[14] P. L. Combettes and B. C. Vũ. Variable metric quasi-Féjer monotonicity. *Nonlinear Anal.*, 78:17–31, feb 2013.

[15] C. Cortes and V. N. Vapnik. Support vector network. *Mach. Learn.*, 20:1–25, 1995.

[16] S. Crisci, V. Ruggiero, and L. Zanni. Steplength selection in gradient projection methods for box-constrained quadratic programs. *Appl. Math. and Comput.*, 356:312–327, 2018.

[17] Y. H. Dai and R. Fletcher. On the asymptotic behaviour of some new gradient methods. *Math. Program.*, 103:541–559, 2005.

[18] Y. H. Dai and R. Fletcher. New Algorithms for Singly Linearly Constrained Quadratic Programming Problems Subject to Lower and Upper Bounds. *Math. Program.*, 106(3):403–421, 2006.

[19] Y. H. Dai, W. H. Hager, K. Schittkowski, and H. Zhang. The cyclic Barzilai-Borwein method for unconstrained optimization. *IMA J. Numer. Anal.*, 26:604–627, 2006.

[20] A. De Asmundis, D. di Serafino, H. Hager, G. Toraldo, and H. Zhang. An efficient gradient method using the Yuan steplength. *Comput. Optim. Appl.*, 59(3):541–563, 2014.

[21] R. De Asmundis, D. di Serafino, F. Riccio, and G. Toraldo. On spectral propertis of steepest descent methods. *IMA J. Numer. Anal.*, 33:1416–1435, 2013.

[22] D. di Serafino, V. Ruggiero, G. Toraldo, and L. Zanni. On the steplength selection in gradient methods for unconstrained optimization. *Appl. Math. Comput.*, 318:176–195, 2018.

[23] E. D. Dolan and J. J. Moré. Benchmarking optimization software with performance profiles. *Math. Program.*, 91(2):201–213, 2002.

[24] F. Facchinei, J. Judice, and J. Soares. Generating box-constrained optimization problems. *ACM Trans. Math. Softw.*, 23(3):443–447, 1997.

[25] R Fletcher. Low storage methods for unconstrained optimization. *Lectures in Appl. Math.*, 26:165–179, 1990.

[26] R. Fletcher. On the Barzilai-Borwein method. In L. Qi, K. Teo, X. Yang, P. M. Pardalos, and D. Hearn, editors, *Optimization and Control with Applications*, volume 96 of *Applied Optimization*, pages 235–256. Springer, US, 2005.

[27] R. Fletcher and M. J. D. Powell. A rapidly convergent descent method for minimization. *Comput. J.*, 6:163–168, 1963.

[28] G. Frassoldati, G. Zanghirati, and L. Zanni. New adaptive stepsize selections in gradient methods. *J. Ind. Manag. Optim.*, 4(2):299–312, 2008.

[29] A. Friedlander, J. M. Martínez, B. Molina, and M. Raydan. Gradient Method with Retards and Generalizations. *SIAM J. on Numer. Anal.*, 36:275–289, 1999.

[30] L. Grippo, F. Lampariello, and S. Lucidi. A nonmonotone line-search technique for Newton's method. *SIAM J. Numer. Anal.*, 23(4):707–716, 1986.

[31] L. Grippo and M. Sciandrone. Nonmonotone globalization techniques for the Barzilai-Borwein gradient method. *Comput. Optim. Appl.*, 23(2):143–169, 2002.

[32] William W. Hager and Hongchao Zhang. An affine scaling method for optimization problems with polyhedral constraints. *Comput. Optim. Appl.*, 59:163–183, 2014.

[33] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, New York, NY, USA, 2nd edition, 2012.

[34] B. Jian and B. Vermuri. A unified computational framework for deconvolution to reconstruct multiple fibers from diffusion weighted MRI. *IEEE Trans. Med. Imag.*, 26:1464–1471, 2007.

[35] H. Lantéri, M. Roche, and C. Aime. Penalized maximum likelihood image restoration with positivity constraints: multiplicative algorithms. *Inverse Problems*, 18:1397–1419, 2002.

[36] H. Lantéri, M. Roche, O. Cuevas, and C. Aime. A general method to devise maximum-likelihood signal restoration multiplicative algorithms with nonnegativity constraints. *Sig. Process.*, 81:945–974, 2001.

[37] I. Loris, M. Bertero, C. De Mol, R. Zanella, and L. Zanni. Accelerating gradient projection methods for ℓ_1-constrained signal recovery by steplength selection rules. *Appl. Comput. Harmon. Anal.*, 27:247–154, 2009.

[38] V. A. Marchenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457–483, 1967.

[39] B. Molina and M. Raydan. Preconditioned Barzilai-Borwein method for the numerical solution of partial differential equations. *Numer. Algorithms*, 13(1):45–60, 1996.

[40] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer–Verlag, New York, Inc., 1999.

[41] P. M. Pardalos and Rosen J. B. *Constrained global optimization: algorithms and applications.* Springer-Verlag, New York, NY, USA, 1987.

[42] F. Porta, R. Zanella, G. Zanghirati, and L. Zanni. Limited-memory scaled gradient projection methods for real-time image deconvolution in microscopy. *Commun. Nonlinear Sci. Numer. Simul.*, 21:112–127, 2015.

[43] T. Serafini, G. Zanghirati, and L. Zanni. Gradient projection methods for large quadratic programs and applications in training support vector machines. *Optim. Methods Softw.*, 20:353–378, 2005.

[44] P. L. Toint. Some numerical results using a sparse matrix updating formula in unconstrained optimization. *Math. Comput.*, 32:839–852, 1978.

[45] V. N. Vapnik and S. Kotz. *Estimation of dependences based on empirical data.* Springer-Verlag, New York, NY, USA, 1982.

[46] A. Zhigljavsky, L. Pronzato, and E. Bukina. An asymptotically optimal gradient algorithm for quadratic optimization with low computational cost. *Optim. Lett., Springer Verlag*, 7(6):1047–1059, 2013.

[47] B. Zhou, L. Gao, and Y. H. Dai. Gradient methods with adaptive step-sizes. *Comput. Optim. Appl.*, 35(1):69–86, 2006.

[48] B. Zhou, L. Gao, and Y. H. Dai. Monotone projected gradient methods for large-scale box-constrained quadratic programming. *Sci. China Ser. A*, 49(5):688–702, 2006.