Steplength selection in gradient projection methods for box-constrained quadratic programs

Serena Crisci[a,c], Valeria Ruggiero[a,c], Luca Zanni[b,c,*]

[a]*Department of Mathematics and Computer Science, University of Ferrara,*
via Saragat 1, I-44122 Ferrara, Italy
[b]*Department of Physics, Informatics and Mathematics,*
University of Modena and Reggio Emilia, via Campi 213/B, I-41125, Modena, Italy
[c]*Member of the INdAM Research group GNCS*

Abstract

The role of the steplength selection strategies in gradient methods has been widely investigated in the last decades. Starting from the work of Barzilai and Borwein (1988), many efficient steplength rules have been designed, that contributed to make the gradient approaches an effective tool for the large-scale optimization problems arising in important real-world applications. Most of these steplength rules have been thought in unconstrained optimization, with the aim of exploiting some second-order information for achieving a fast annihilation of the gradient of the objective function. However, these rules are successfully used also within gradient projection methods for constrained optimization, though, to our knowledge, a detailed analysis of the effects of the constraints on the steplength selections is still not available. In this work we investigate how the presence of the box constraints affects the spectral properties of the Barzilai-Borwein rules in quadratic programming problems. The proposed analysis suggests the introduction of new steplength selection strategies specifically designed for taking account of the active constraints at each iteration. The results of a set of numerical experiments show the effectiveness of the new rules with respect to other state of the art steplength selections and their potential usefulness also in case of box-constrained non-quadratic optimization problems.

Keywords: box-constrained optimization, gradient projection methods, steplength rules, Hessian spectral properties.
2010 MSC: 65K05, 90C20, 90C30, 49M05

1. Introduction

Gradient methods are widely used for solving nonlinear optimization problems and, in many large-scale applications, their simplicity and low memory requirements make them the most convenient choice. In the last years, very efficient gradient-based approaches have been designed for unconstrained and constrained problems arising in different areas, such as image processing, compressive sensing and machine learning. The success of the gradient schemes

---

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

in these applications is mainly due to the introduction of special strategies for accelerating their convergence rate, such as the adaptive techniques for the steplength selection, the use of extrapolation steps or the employment of scaled gradient directions.

This work deals with steplength selection rules, by investigating how popular techniques designed in the context of unconstrained optimization can be adapted to gradient projection methods for box-constrained quadratic programming problems. We focus on state of the art steplength strategies designed by exploiting the well-known Barzilai-Borwein (BB) rules [1]. These strategies have been originally developed in the unconstrained quadratic case, by taking into account the importance of using steplengths approximating the inverse of the eigenvalues of the Hessian matrix for achieving fast reductions of the eigencomponents of the gradient. In [2] the spectral properties of both the standard BB rule and other improved steplength techniques [3, 4, 5, 6] have been investigated. It has been shown that the steplength rules specifically designed for an efficient approximation of the spectrum of the Hessian generally outperform the standard BB rule and are preferable also when applied for providing the tentative steplength of a line search procedure within gradient methods for general non-quadratic unconstrained problems [7, 8, 9, 10, 11].

For what concerns the case of the constrained optimization problems, the role of the steplength rules is not so deeply investigated. Many well-known gradient projection methods for constrained optimization simply exploit the same steplength selections designed for the unconstrained case in combination with some kind of line search strategy [12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22]. However, most of the steplength selections for unconstrained problems are designed for achieving a fast annihilation of the gradient of the objective function, that is not the main goal in constrained optimization. Thus, a better understanding of the role of the steplength in gradient projection methods seems necessary and useful for improving first-order approaches in constrained optimization.

In this paper, for the case of box-constrained quadratic programming problems, we analyze how the two popular BB steplength rules behave within gradient projection methods and how their spectral properties are affected by the presence of the constraints. In particular, on the basis of this spectral analysis, we design a new steplength rule that, by taking into account the active constraints at each iteration, is able to achieve fast reductions of the gradient components corresponding to the free variables at the solution. Numerical experiments on different sets of randomly generated test problems are carried out for evaluating the practical effectiveness of the gradient projection methods exploiting steplength selections based on the proposed rule. Preliminary numerical results on box-constrained non-quadratic problems are also discussed for showing the potential usefulness of the new steplength rule in this more general framework.

The outline of the paper is the following. In Section 2, after briefly recalling the gradient projection methods and the steplength rules we are interested to, we provide a spectral analysis of the standard BB rules in the case of box-constrained quadratic programs and introduce the new steplength selection; futhermore, alternation steplength strategies based on the new rule are discussed. In Section 3 we report the results of the numerical experiments and, finally, some conclusions are drawn in Section 4.

Algorithm 1 GP method for box-constrained quadratic programs

Initialize: Choose $x^{(0)} \in \mathbb{R}^n$, $\ell \le x^{(0)} \le u$, $\delta, \sigma \in (0,1)$, $M \in \mathbb{N}$, $0 < \alpha_{min} \le \alpha_{max}$,
$\alpha_0 \in [\alpha_{min}, \alpha_{max}]$;

for $k = 0, 1, \ldots$

$d^{(k)} = P_{\ell \le x \le u}\left(x^{(k)} - \alpha_k g(x^{(k)})\right) - x^{(k)};$ (gradient projection step)

$\nu_k = 1;$ $f_{ref} = \max\{f(x^{(k-i)}),\ 0 \le i \le \min(k, M)\};$

while $f(x^{(k)} + \nu_k d^{(k)}) > f_{ref} + \sigma \nu_k g(x^{(k)})^T d^{(k)}$

   $\nu_k = \delta \nu_k;$ (backtracking step)

end

$x^{(k+1)} = x^{(k)} + \nu_k d^{(k)};$

define the steplength $\alpha_{k+1} \in [\alpha_{min}, \alpha_{max}];$ (steplength updating rule)

end

---

2. Steplength rules in GP methods for box-constrained quadratic programs

We consider the following box-constrained Quadratic Programming (QP) problem

$$\min_{\ell \le x \le u} f(x) \equiv \frac{1}{2} x^T A x - b^T x + c \tag{1}$$

where A is an $n \times n$ symmetric positive definite matrix, b, ℓ, u are vectors of $\mathbb{R}^n$, with $\ell \le u$, and c is a scalar. Problem (1) has a unique solution x^*, satisfying the following first-order optimality conditions

$$\begin{cases} g(x^*)_i = 0 & \text{for} \quad \ell_i < x_i^* < u_i, \\ g(x^*)_i \le 0 & \text{for} \quad x_i^* = u_i, \\ g(x^*)_i \ge 0 & \text{for} \quad x_i^* = \ell_i, \end{cases} \tag{2}$$

where $g(x)$ denotes the gradient of the objective function at x: $g(x) \equiv \nabla f(x) = Ax - b$.

In this work we are interested in solving problem (1) by means of the well-known Gradient Projection (GP) method combined with a line search strategy along the feasible direction. The main steps of the GP method are described in Algorithm 1. At each iteration, the operator $P_{\ell \le x \le u}(\cdot)$ is exploited for computing the Euclidean projection of the vector $\left(x^{(k)} - \alpha_k g(x^{(k)})\right)$ onto the feasible region given by the box constraints, in order to obtain the direction $d^{(k)}$, which is a feasible descent direction for the objective function at $x^{(k)}$. The new iterate is defined as

$$x^{(k+1)} = x^{(k)} + \nu_k d^{(k)}, \tag{3}$$

where $0 < \nu_k \le 1$ is determined by a backtracking procedure implementing a standard nonmonotone line search [23]. Finally, a steplength updating rule sets the parameter $\alpha_{k+1} \in [\alpha_{min}, \alpha_{max}]$ for the next iteration. Our analysis focuses just on the steplength updating strategies and their relationship with the special constraints of problem (1). Before deepening this topic, we recall in Proposition 1 the basic convergence results for the sequence $\{x^{(k)}\}$ generated by Algorithm 1.

Proposition 1. *Assume that the matrix A in problem (1) is symmetric and positive definite, then the following properties hold for the sequence $\{x^{(k)}\}$ generated by Algorithm 1:*

(i) the sequence $\{x^{(k)}\}$ converges R-linearly to the unique solution x^* of problem (1);

3

(ii) the sequence $\{f(x^{(k)})\}$ has $\mathcal{O}\left(\frac{1}{k}\right)$ convergence rate.

The proof of Proposition 1 can be derived from the analysis developed in [24] for the case in which the GP method is applied to the minimization of general strongly convex functions. For basic properties of the method we refer, for instance, to [12, 25] and for a convergence analysis of more general scaled GP methods to [13, 26, 27].

It is worth stressing that the convergence properties in Proposition 1 hold independently of the choice of the steplength α_k in the closed interval $[\alpha_{min}, \alpha_{max}]$. This is very important from a practical point of view, since it allows one to make the updating rule of α_k oriented at optimizing the performance. Many evidences of the remarkable convergence rate improvements achievable by means of suitable steplength updating rules are available in literature; in particular, promising results have been obtained, both on library test problems and real-world applications, by exploiting the Barzilai-Borwein steplength rules [12, 20, 26, 18, 14, 15, 17]. These rules, originally introduced for unconstrained minimization problems [1], aim at capturing some second order information by forcing the matrix $(\alpha_k I_n)^{-1}$ to approximate the Hessian of the objective function through one of the following secant conditions:

$$\alpha_k = \arg\min_\alpha \left\| \alpha^{-1} s^{(k-1)} - y^{(k-1)} \right\| \tag{4}$$

or

$$\alpha_k = \arg\min_\alpha \left\| s^{(k-1)} - \alpha y^{(k-1)} \right\|, \tag{5}$$

where $s^{(k-1)} = x^{(k)} - x^{(k-1)}$ and $y^{(k-1)} = g^{(k)} - g^{(k-1)}$, where the notation $g^{(i)} = g(x^{(i)})$ is exploited. From (4) and (5) the steplength updating rules

$$\alpha_k^{\mathrm{BB1}} = \frac{\| s^{(k-1)} \|^2}{(s^{(k-1)})^T y^{(k-1)}} \tag{6}$$

and

$$\alpha_k^{\mathrm{BB2}} = \frac{(s^{(k-1)})^T y^{(k-1)}}{\| y^{(k-1)} \|^2} \tag{7}$$

are obtained, respectively. In the last years a lot of effort has been put into understanding the reasons of the effectiveness of the BB rules and great attention has been devoted to their ability in approximating the inverse of the eigenvalues of the objective Hessian. In fact, in the case of the unconstrained minimization of the quadratic function in (1), a gradient method is simply described by the iteration

$$x^{(k+1)} = x^{(k)} - \alpha_k g^{(k)}, \tag{8}$$

and for the gradient $g^{(k+1)}$ we have

$$g^{(k+1)} = g^{(k)} - \alpha_k A g^{(k)}. \tag{9}$$

By denoting with $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ the eigenvalues of A and with $\{v_1, v_2, \ldots, v_n\}$ a set of associated orthonormal eigenvectors, the gradient $g^{(k+1)}$ can be expressed as

$$g^{(k+1)} = \sum_{i=1}^n \mu_i^{(k+1)} v_i, \tag{10}$$

4

where $\mu_i^{(k+1)} \in \mathbb{R}$ is called the i-th eigencomponent of $g^{(k+1)}$ and satisfies the following recurrence formula:

$$\mu_i^{(k+1)} = \mu_i^{(0)} \prod_{j=0}^{k}(1 - \alpha_j \lambda_i) = \mu_i^{(k)}(1 - \alpha_k \lambda_i). \tag{11}$$

The previous recurrence highlights the crucial role of a steplength α_k approximating an inverse of an eigenvalue λ_i: it forces a remarkable reduction of $|\mu_i^{(k+1)}|$ with respect to $|\mu_i^{(k)}|$, even if it can amplify the absolute value of some other eigencomponents. Thus, the behaviour of a steplength selection rule can be fruitfully studied in terms of its ability in reducing all the gradient eigencomponents by means of an effective sweeping of the spectrum of A^{-1}. When the gradient iteration (8) is applied in unconstrained quadratic optimization, the BB rules (6) and (7) can be written as

$$\alpha_k^{\mathrm{BB1}} = \frac{g^{(k-1)^T}g^{(k-1)}}{g^{(k-1)^T}Ag^{(k-1)}}, \tag{12}$$

$$\alpha_k^{\mathrm{BB2}} = \frac{g^{(k-1)^T}Ag^{(k-1)}}{g^{(k-1)^T}A^2g^{(k-1)}}. \tag{13}$$

These steplength rules are suitable for providing approximations of the eigenvalues of A^{-1}, since they satisfy

$$\frac{1}{\lambda_{\max}(A)} \leq \alpha_k^{\mathrm{BB2}} \leq \alpha_k^{\mathrm{BB1}} \leq \frac{1}{\lambda_{\min}(A)}, \tag{14}$$

where $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the minimum and the maximum eigenvalues of A, respectively. In [2], a spectral analysis of popular steplength selections shows that the BB rule (12) produces steplengths within the spectrum of A^{-1} in an almost random way, while other techniques, which are aimed at exploiting groups of small steplengths followed by some large steplengths [3, 4, 5, 28, 29], seem to better adapt to the spectrum of A^{-1}, thanks to their skill in catching the eigenvalues in a suitable order. In particular, very promising convergence rate improvements with respect to the standard BB rules have been provided by the Adaptive Barzilai-Borwein (ABB) strategy [3] and its modification $\mathrm{ABB}_{\min}$ [4], that are defined by

$$\alpha_k^{\mathrm{ABB}} = \begin{cases} \alpha_k^{\mathrm{BB2}} & \text{if } \dfrac{\alpha_k^{\mathrm{BB2}}}{\alpha_k^{\mathrm{BB1}}} < \tau \\ \alpha_k^{\mathrm{BB1}} & \text{otherwise} \end{cases} \tag{15}$$

and

$$\alpha_k^{\mathrm{ABB_{\min}}} = \begin{cases} \min\left\{\alpha_j^{\mathrm{BB2}} \, : \, j = \max\{1, k - m_a\}, \dots, k\right\} & \text{if } \dfrac{\alpha_k^{\mathrm{BB2}}}{\alpha_k^{\mathrm{BB1}}} < \tau \\ \alpha_k^{\mathrm{BB1}} & \text{otherwise} \end{cases} \tag{16}$$

respectively, where $\tau \in (0,1)$ and m_a is a nonnegative integer.

The BB rules (6), (7) and their adaptive alternations (15) and (16) can be exploited also in gradient methods for general nonlinear optimization problems; in these cases, they show a behaviour very similar to that observed for quadratic problems, although the recurrence

relation (9) does not hold. In particular, in [2] it is shown that the adaptive alternation (16) appears more efficient than the rule (6) in fast approximating the spectrum of the Hessian matrices of the objective function and, consequently, shows a better convergence rate.

Even if all these BB-based steplength rules have been designed in the framework of unconstrained optimization, they are widely used in gradient projection methods for constrained minimization problems, still providing interesting performances [12, 26, 14, 16, 19, 15]. Here, we investigate how the spectral analysis proposed in [2] can be generalized to the GP method for problem (1), with the aim to provide some explanations of the BB effectiveness in this special constrained optimization setting. We observe that, in this case, neither the recurrence formula (9) nor the equation (11) hold, and the optimality conditions (2) require that only some gradient components annihilate at the solution. These remarkable differences with respect to the unconstrained case are at the basis of the study developed in the next subsection.

2.1. BB-like rules for box-constrained quadratic programs

In order to understand the spectral properties of the BB rules (6) and (7) when they are exploited within the GP method for solving problem (1), we need to investigate how the gradient projection step influences their definitions. We first study the BB1 rule and then, by proceeding in a similar way, we develop the analysis for the BB2 selection. If, at the $(k-1)$-th iteration, we suppose that the j-th component of $x^{(k-1)}$ is active, $x_j^{(k-1)} = \ell_j$ (or $x_j^{(k-1)} = u_j$), and the corresponding gradient component is non-negative (or non-positive), then we have $d_j^{(k-1)} = 0$, so that $x_j^{(k)} = x_j^{(k-1)}$. As a consequence, the corresponding entry of the vector $s^{(k-1)}$ used in the steplength definition (6) is equal to zero: $s_j^{(k-1)} = 0$. By using the notation

$$\mathcal{J}_{k-1} = \{i: \ (x_i^{(k-1)} = \ell_i \ \wedge \ g_i^{(k-1)} \geq 0) \ \vee \ (x_i^{(k-1)} = u_i \ \wedge \ g_i^{(k-1)} \leq 0)\} \qquad (17)$$

and $\mathcal{I}_{k-1} = \{1, ..., n\} \setminus \mathcal{J}_{k-1}$, we can write

$$\alpha_k^{\text{BB1}} = \frac{\|s^{(k-1)}\|^2}{(s^{(k-1)})^T y^{(k-1)}} = \frac{\|s_{\mathcal{I}_{k-1}}^{(k-1)}\|^2}{\left(s_{\mathcal{I}_{k-1}}^{(k-1)}\right)^T y_{\mathcal{I}_{k-1}}^{(k-1)}}. \qquad (18)$$

This means that when the updating rule (6) is used in Algorithm 1 and $\alpha_k^{\text{BB1}} \in [\alpha_{min}, \alpha_{max}]$, the steplength α_k satisfies the secant condition

$$\alpha_k = \arg \min_\alpha \left\| \alpha^{-1} s_{\mathcal{I}_{k-1}}^{(k-1)} - y_{\mathcal{I}_{k-1}}^{(k-1)} \right\|^2, \qquad (19)$$

that is, it is defined by forcing the matrix $(\alpha_k I)^{-1}$ to approximate the submatrix $A_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}}$ of A defined by the intersection of the rows and the columns with indices in $\mathcal{I}_{k-1}$. We call this submatrix the *reduced Hessian matrix* at the $(k-1)$-th iteration.

In particular, in the following theorem we prove that the inverse of α_k^{BB1} belongs to the spectrum of $A_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}}$.

Theorem 1. *If A in (1) is a symmetric positive definite matrix, we have*

$$\lambda_{min}(A_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}}) \leq \frac{1}{\alpha_k^{\text{BB1}}} \leq \lambda_{max}(A_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}}), \qquad (20)$$

6

where $\lambda_{min}(A_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}})$ and $\lambda_{max}(A_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}})$ are the minimum and the maximum eigenvalues of $A_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}}$, respectively.

Proof: We assume that at the iteration $(k-1)$ the rows/columns of A and the entries of any vector are reordered so that $\mathcal{I}_{k-1}$ is related to the first indices and $\mathcal{J}_{k-1}$ contains the last indices; by dropping for simplicity the iteration counter $k-1$ from $\mathcal{I}_{k-1}$ and $\mathcal{J}_{k-1}$, we can write

$$x^{(k-1)} = \begin{pmatrix} x_{\mathcal{I}}^{(k-1)} \\ x_{\mathcal{J}}^{(k-1)} \end{pmatrix}, \qquad A = \begin{pmatrix} A_{\mathcal{I},\mathcal{I}} & A_{\mathcal{I},\mathcal{J}} \\ A_{\mathcal{I},\mathcal{J}}^T & A_{\mathcal{J},\mathcal{J}} \end{pmatrix}. \tag{21}$$

In view of the GP iteration, we have that the entries of the iterate $x^{(k)}$ are

$$x_i^{(k)} = \begin{cases} x_i^{(k-1)} + \nu_{k-1}(p_i^{(k-1)} - x_i^{(k-1)}) & \text{for } i \in \mathcal{I} \\ x_i^{(k-1)} & \text{for } i \in \mathcal{J} \end{cases} \tag{22}$$

where $p_i^{(k-1)} = \max(\ell_i, \min(x_i^{(k-1)} - \alpha_{k-1} g_i^{(k-1)}, u_i)), i \in \mathcal{I}$. By denoting with $p^{(k-1)}$ the vector of length $\sharp\mathcal{I}$ with entries $p_i^{(k-1)}, i \in \mathcal{I}$, the vector $s^{(k-1)}$ can be partitioned into two sub-vectors given by

$$s^{(k-1)} = \begin{pmatrix} s_{\mathcal{I}}^{(k-1)} \\ s_{\mathcal{J}}^{(k-1)} \end{pmatrix} = \begin{pmatrix} \nu_{k-1}(p^{(k-1)} - x_{\mathcal{I}}^{(k-1)}) \\ 0 \end{pmatrix}. \tag{23}$$

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

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

Consequently, from (23), we can write

$$y^{(k-1)} = \begin{pmatrix} y_{\mathcal{I}}^{(k-1)} \\ y_{\mathcal{J}}^{(k-1)} \end{pmatrix} = g^{(k)} - g^{(k-1)} = \begin{pmatrix} A_{\mathcal{I},\mathcal{I}} s_{\mathcal{I}}^{(k-1)} \\ A_{\mathcal{J},\mathcal{I}} s_{\mathcal{I}}^{(k-1)} \end{pmatrix}. \tag{24}$$

From the definition (18), the value of α_k^{BB1} can be written as

$$\alpha_k^{\text{BB1}} = \frac{(s_{\mathcal{I}}^{(k-1)})^T s_{\mathcal{I}}^{(k-1)}}{(s_{\mathcal{I}}^{(k-1)})^T A_{\mathcal{I},\mathcal{I}} s_{\mathcal{I}}^{(k-1)}}, \tag{25}$$

or, in other words, α_k^{BB1} is the reciprocal of a Rayleigh quotient of the submatrix $A_{\mathcal{I},\mathcal{I}}$. $\square$

Theorem 1 highlights that the use of the BB1 rule in Algorithm 1 leads to capture spectral properties of the matrix $A_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}}$ instead of the whole matrix A, as done in the unconstrained quadratic case. Thus, we may argue that now the BB1 rule plays a role in

reducing the gradient components with indices belonging to $\mathcal{I}_{k-1}$. This can be observed in the special case in which $\mathcal{I}_k = \mathcal{I}_{k-1}$ and $p_i^{(k)} = (x_i^{(k)} - \alpha_k g_i^{(k)})$, $i \in \mathcal{I}_k$. In fact, under these assumptions we have

$$g_{\mathcal{I}_k}^{(k+1)} = g_{\mathcal{I}_k}^{(k)} + A_{\mathcal{I}_k,\mathcal{I}_k} s_{\mathcal{I}_k}^{(k)} = g_{\mathcal{I}_k}^{(k)} - \nu_k \alpha_k A_{\mathcal{I}_k,\mathcal{I}_k} g_{\mathcal{I}_k}^{(k)}.$$

By denoting with $\gamma_1, \ldots, \gamma_r$ and $u_1, \ldots, u_r$ the eigenvalues and the associated orthonormal eigenvectors of $A_{\mathcal{I}_k,\mathcal{I}_k}$, respectively, where $r = \sharp \mathcal{I}_k$, and by writing $g_{\mathcal{I}_k}^{(k+1)} = \sum_{i=1}^r \bar{\mu}_i^{(k+1)} u_i$ and $g_{\mathcal{I}_k}^{(k)} = \sum_{i=1}^r \bar{\mu}_i^{(k)} u_i$, we obtain the following recurrence formula for the eigencomponents:

$$\bar{\mu}_i^{(k+1)} = \bar{\mu}_i^{(k)}(1 - \nu_k \alpha_k \gamma_i).$$

This means that if the selection rule (25) provides a good approximation of $\frac{1}{\gamma_i}$, a useful reduction of $|\bar{\mu}_i^{(k+1)}|$ can be achieved.

With regard to the second BB rule, using the above notation for the $(k-1)$-th iteration of GP, we obtain that the steplength (7) can be written as

$$\alpha_k^{\text{BB2}} = \frac{(s_{\mathcal{I}_{k-1}}^{(k-1)})^T y_{\mathcal{I}_{k-1}}^{(k-1)}}{\|y_{\mathcal{I}_{k-1}}^{(k-1)}\|^2 + \|y_{\mathcal{J}_{k-1}}^{(k-1)}\|^2}. \tag{26}$$

In view of (24), we have

$$\alpha_k^{\text{BB2}} = \frac{(s_{\mathcal{I}_{k-1}}^{(k-1)})^T A_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}} s_{\mathcal{I}_{k-1}}^{(k-1)}}{(s_{\mathcal{I}_{k-1}}^{(k-1)})^T A_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}}^2 s_{\mathcal{I}_{k-1}}^{(k-1)} + (s_{\mathcal{I}_{k-1}}^{(k-1)})^T A_{\mathcal{J}_{k-1},\mathcal{I}_{k-1}}^T A_{\mathcal{J}_{k-1},\mathcal{I}_{k-1}} s_{\mathcal{I}_{k-1}}^{(k-1)}} \tag{27}$$

and, consequently, $1/\alpha_k^{\text{BB2}}$ might be outside of the spectrum of the current reduced Hessian at $x^{(k-1)}$. A simple way to correct (26) is to redefine this value as

$$\alpha_k^{\text{BoxBB2}} = \frac{(s_{\mathcal{I}_{k-1}}^{(k-1)})^T y_{\mathcal{I}_{k-1}}^{(k-1)}}{\|y_{\mathcal{I}_{k-1}}^{(k-1)}\|^2}, \tag{28}$$

so that the following result holds.

Theorem 2. *If A in (1) is a symmetric positive definite matrix, we have*

$$\lambda_{min}(A_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}}) \leq \frac{1}{\alpha_k^{\text{BoxBB2}}} \leq \lambda_{max}(A_{\mathcal{I}_{k-1},\mathcal{I}_{k-1}}). \tag{29}$$

The proof of Theorem 2 follows immediately by using the definition (24) of $y_{\mathcal{I}_{k-1}}^{(k-1)}$.

From the comparison between (27) and (28) we have that $\alpha_k^{\text{BB2}} \leq \alpha_k^{\text{BoxBB2}}$; furthermore, in analogy with the inequalities (14), for the rules α_k^{BB1} and α_k^{BoxBB2} we have:

$$\frac{1}{\lambda_{max}(A_{\mathcal{I}_{k-1}\mathcal{I}_{k-1}})} \leq \alpha_k^{\text{BoxBB2}} \leq \alpha_k^{\text{BB1}} \leq \frac{1}{\lambda_{min}(A_{\mathcal{I}_{k-1}\mathcal{I}_{k-1}})}. \tag{30}$$

In view of these inequalities, the modified BB2 rule (28) can be exploited within the adaptive strategies (15) and (16) that use the BB steplengths in an alternate way, for an effective combination of short and long steps also in the framework of gradient methods for constrained

optimization. In the following, we denote by $\text{BoxABB}_{\min}$ the modified $\text{ABB}_{\min}$ selection in which the BB2 rule is substituted by the BoxBB2 defined in (28):

$$\alpha_k^{\text{BoxABB}_{\min}} = \begin{cases} \min\left\{\alpha_j^{\text{BoxBB2}} : j = \max\{1, k - m_a\}, \dots, k\right\} & \text{if } \dfrac{\alpha_k^{\text{BoxBB2}}}{\alpha_k^{\text{BB1}}} < \tau \\ \alpha_k^{\text{BB1}} & \text{otherwise} \end{cases} \quad (31)$$

where $\tau \in (0, 1)$ and m_a is a nonnegative integer; furthermore, we denote by $\text{BoxVABB}_{\min}$ the $\text{BoxABB}_{\min}$ variant in which a variable setting of the parameter τ is exploited, as suggested in [26]:

$$\tau_{k+1} = \begin{cases} \tau_k / \vartheta & \text{if } \dfrac{\alpha_k^{\text{BoxBB2}}}{\alpha_k^{\text{BB1}}} < \tau_k \\ \tau_k \cdot \vartheta & \text{otherwise} \end{cases} \quad (32)$$

with $\vartheta > 1$. A typical value for ϑ is 1.1. It is worthwhile observing that this variable setting makes the efficiency of the steplength strategy less dependent on the value of τ provided by the user and in several applications it allowed remarkable performance improvements with respect to the standard $\text{ABB}_{\min}$ strategy [26, 19, 30, 31].

In the next section, we will evaluate on box-constrained test problems the behaviour of the proposed modified steplengths within the GP method. In particular, we will focus on the ability of these steplength rules to capture the spectral properties of the reduced Hessian, recovering in this way the benefits usually exhibited by the standard BB rules in case of unconstrained problems.

3. Numerical Experiments

In order to highlight the effectiveness of the modified versions of the BB rules for box-constrained QP problems, we perform two numerical investigations: in the former we analyze the spectral behaviour of the different steplength rules on some toy problems, in the latter we evaluate the rules on a set of large-scale benchmark test problems. Finally, in order to have some hints about the behaviour of the modified BB rules for box-constrained non-quadratic minimization problems, we report the numerical results obtained by the GP method combined with these rules on some well-known general test problems.
All the numerical experiments are carried out on a workstation equipped with an Intel Xeon QuadCore E5620 processor at 2,40 GHz and 18 Gb of RAM, by implementing the GP methods in the Matlab R2016a environment.

3.1. Numerical results on some special box-constrained QP problems

In this subsection we study the distribution of the steplengths generated by the considered rules with respect to the eigenvalues of the reduced Hessian matrices obtained during the GP iterative process. To this end, we randomly generated QP problems subject to lower bounds in which the solution, the number of active constraints at the solution and the distribution of the eigenvalues of the dense symmetric positive definite Hessian of the objective function are prefixed. The main features of the three test problems used for this analysis are described in Table 1, where na is the number of active lower constraints at the solution and $\mathcal{I}^*$ denotes the set of the indices of the inactive constraints at the solution, so that $A_{\mathcal{I}^*, \mathcal{I}^*}$ is the reduced Hessian matrix at the solution. The GP variants exploiting the different steplength rules are

Table 1: Main features of three QP test problems subject to lower bounds

	$\lambda_{min}(A)$	$\lambda_{max}(A)$	$n = 1000 \qquad na = 400$ distribution of the eigenvalues of A	$\lambda_{min}(A_{\mathcal{I}^*,\mathcal{I}^*})$	$\lambda_{max}(A_{\mathcal{I}^*,\mathcal{I}^*})$
TP1	1	1000	$\lambda_i(A) = \frac{1}{2}\left(\overline{\lambda} + \underline{\lambda} + (\overline{\lambda} - \underline{\lambda})cos\left(\frac{\pi(i-1)}{n-1}\right)\right)$ $\underline{\lambda} = 1 \qquad \overline{\lambda} = 1000$	41.78	957.19
TP2	1	1000	log-spaced	3.08	753.26
TP3	9.40	1013.95	log-spaced	10	1000

distinguished by means of the rule's name (BB1, BB2, BoxBB2, BoxABB$_{min}$, BoxVABB$_{min}$) and share the same stopping criterion:

$$\|\varphi(x^{(k)})\| \leq tol\|g(x^{(0)})\|, \qquad (33)$$

where $\varphi(x^{(k)})$ is projected gradient at $x^{(k)}$, i.e., the vector with entries $\varphi_i^{(k)}$, $i = 1 \ldots, n$, defined as

$$\varphi_i^{(k)} = \begin{cases} g_i^{(k)} & \text{for} \quad \ell_i < x_i^{(k)} < u_i, \\ \max\{0, g_i^{(k)}\} & \text{for} \quad x_i^{(k)} = u_i, \\ \min\{0, g_i^{(k)}\} & \text{for} \quad x_i^{(k)} = \ell_i. \end{cases} \qquad (34)$$

In the methods the following parameter setting is used: $tol = 10^{-8}$, $M = 9$, $\sigma = 10^{-4}$, $\delta = 0.5$, $\alpha_{min} = 10^{-10}$, $\alpha_{max} = 10^6$, $\alpha_0 = (g^{(0)T}g^{(0)})/(g^{(0)T}Ag^{(0)})$; furthermore, $\tau = \tau_1 = 0.5$ and $m_\alpha = 2$ are set in BoxABB$_{min}$ and BoxVABB$_{min}$. The feasible initial point $x^{(0)}$ is randomly generated such that its entries are inactive.

Figures 1-3 enable to show the behaviour of the BB1, BB2, BoxBB2 and BoxABB$_{min}$ rules on the three test problems. In the top and middle panels of the Figures, at the k-th iteration, the black dots denote 20 eigenvalues of the reduced Hessian matrix $A_{\mathcal{I}_{k-1}\mathcal{I}_{k-1}}$, with linearly spaced indices (included the maximum and the minimum eigenvalues), and the red cross corresponds to the inverse of the steplength α_k. The blue lines show the maximum and the minimum eigenvalues of the whole Hessian matrix and the blue symbols "o" are used to denote 20 eigenvalues of the reduced Hessian at the prefixed solution x^*, with linearly spaced indices (included the maximum and the minimum eigenvalues). In the bottom panels, the errors $\|x^{(k)} - x^*\|/\|x^*\|$ (left panel) and $f(x^{(k)}) - f(x^*)$ (right panel) corresponding to the different steplength rules are shown. The plots in Figures 1-3 confirm the ability of the BB1 and BoxBB2 rules to produce steplengths α_k whose inverse belong to the spectrum of the reduced Hessian $A_{\mathcal{I}_{k-1}\mathcal{I}_{k-1}}$. Furthermore, in the last iterations the eigenvalues of the reduced Hessian matrices tend to stabilize and to well approximate the eigenvalues of $A_{\mathcal{I}^*\mathcal{I}^*}$. This means that the two rules play a role in reducing the gradient components with indices belonging to I^*, that are just the gradient components annihilated at the solution. On the other hand, the inverses of the BB2 steplengths seem unable to effectively sweep the spectrum of the reduced Hessian matrices with dangerous effects on the performance. The behaviour of the BoxABB$_{min}$ rule suggests that the benefits exhibited by the special adaptive alternation of small and large steplengths introduced by ABB$_{min}$ for the unconstrained case, can be a useful strategy also in the box-constrained case, provided that the BoxBB2 selection is exploited in place of the standard BB2 rule.

3.2. Numerical results on large-scale box-constrained QP problems

In order to verify the above conclusions on a more exhaustive set of large-scale problems, we used the software package downloadable at `http://www.dimat.unina2.it/diserafino/dds_sw.htm` for randomly generating box-constrained quadratic problems. Following the procedure proposed in [32, 33], the software allows to generate test problems with different size, spectral properties and number of active constraints at the solution. We generated a dataset of 108 strictly convex QP problems with nondegenerate solutions, splitted into three groups of increasing size: $n = 15000, 20000, 25000$; for each group, 36 problems are generated by considering three values for the number na of active constraints at the solution, $na \approx 0.1 \cdot n, 0.5 \cdot n, 0.9 \cdot n$, three values for the condition number $\kappa(A)$ of the Hessian matrix A, $\kappa(A) = 10^4, 10^5, 10^6$, and four levels of near-degeneracy, obtained by setting the positive Lagrangian multipliers β_i^*, $i = 1, \ldots, na$, equal to $10^{-\eta_i ndeg}$, where $\eta_i \in (0, 1)$ is a random number and $ndeg = 1, 4, 7, 10$. The solution x^* of each problem is randomly chosen from a uniform distribution in $(-1, 1)$, the starting point $x^{(0)}$ is set equal to $\frac{\ell + u}{2}$ and the Hessian matrix is defined as $A = GDG^T$ where G is the product of three Householder matrices associated to randomly generated unit vectors and D is the diagonal matrix of the eigenvalues, which are log-spaced between 1 and $\kappa(A)$. Hence, the matrix A is not stored in memory but only the operator for computing the matrix-vector products involving A is available. The parameter setting of the methods is the same of the previous section. We remark that we did not particularly focus on the optimal choice of the parameter m_α in the steplength rule (31), since in these experiments we observed that, as long as m_a is selected as a small integer, it slightly affects the performance of the BoxABB$_{\min}$ based methods. Furthermore, we remark that, in order to preserve the property that the inverse of the steplength is in the spectrum of the current reduced Hessian also at the initial iterations, it must to choose a small value for m_a, avoiding that a steplength computed by the $\alpha_k^{\mathrm{BoxBB2}}$ rule at an iteration too early with respect the current one is used. The behaviour of the considered steplength rules on these test problems is evaluated by using the performance profiles proposed in [34]. By denoting with "solver" the GP method equipped with a generic steplength rule, we assume as performance measure of interest the execution time required by each solver to satisfy the stopping criterion defined by (33), with $tol = 10^{-7}$. As in [34], given a set $\mathcal{P}$ of n_p problems, for each problem p we consider the ratio $r_{p,s}$ of the computing time of a solver s versus the best time of all the solvers (performance ratio) and, for each solver s and for $\theta \geq 1$ define

$$P(r_{p,s}, \theta) = \begin{cases} 1 & \text{if } r_{p,s} \leq \theta, \\ 0 & \text{otherwise.} \end{cases}$$

The performance profile of a solver s is given by

$$\rho_s(\theta) = \frac{\sum_{p \in \mathcal{P}} P(r_{p,s}, \theta)}{n_p};$$

hence, $\rho_s(\theta)$ is the probability for solver s that $r_{p,s} \leq \theta$. If the stopping criterion is not fulfilled by a solver within the maximum number of 40000 iterations, the corresponding performance ratio is set equal to a fixed value r_M larger than any ratio $r_{p,s}$. For the test problems described above, the performance profiles of the different GP versions are reported in Figure 4. The BoxBB2 steplength rule is compared with the standard BB2 and BB1 rules in the top-left and top-right panels, respectively, while a comparison between the BoxBB2

and the BoxABB$_{\min}$ is provided in the bottom-left panel; a summarizing analysis including also the BoxVABB$_{\min}$ is given in the bottom-right panel. Figure 4 shows that, on the considered set of test problems, the proposed BoxBB2 steplength rule is generally preferable to the standard BB2 and BB1 selections. Furthermore, the adaptive alternation of the BB1 and BoxBB2 selections exploited in the BoxABB$_{\min}$ is convenient with respect to the use of a single steplength rule. In particular, the version BoxVABB$_{\min}$ is able to further improve the performance provided by the BoxABB$_{\min}$ thanks to the updating formula (32) for the switching parameter τ.

Figure 5 shows the performance profiles obtained by the GP method equipped with different steplength rules on the subsets of box-constrained QP test problems generated by setting $ndeg = 7$ (left panel) and $ndeg = 10$ (right panel). The comparison of Figure 5 with the bottom-right panel of Figure 4 makes in evidence that the level of degeneracy of the constraints weakly affects the performance of the different rules. All the above experiments highlight that the rationale behind the BoxBB2 design, aiming at making the rule able to generate values sweeping the spectrum of the inverse of the reduced Hessian matrices, has given rise to a steplength rule providing a promising practical behaviour, both in comparison with the widely used BB1 rule and, in particular, within efficient adaptive alternation strategies, like BoxABB$_{\min}$ and BoxVABB$_{\min}$. Furthermore, we used the same parameter setting to generate two datasets of strictly convex QP problems with degenerate solutions obtained by setting the fraction of active variables at the solution x^* that are degenerate approximately equal to $0.2 \cdot na$ and $0.5 \cdot na$, respectively. The results attained on these datasets are summarized in the performance profiles of Figure 6; the presence of a degenerate solution for these sets of problems does not affect the performance of the steplength rules observed for the dataset of QP problems with a non-degenerate solution.

3.3. Beyond the quadratic case

In case of unconstrained optimization problems, the numerical behaviour exhibited by the BB-like rules in the quadratic case is also observed in case of non-quadratic problems [2]: the steplength selection strategies designed for efficiently adapting to the spectrum of the Hessian matrices of the objective function generally provide the best performance in terms of convergence rate of the gradient scheme. Then, it is natural to ask if in the case of box-constrained non-quadratic problems the proposed BoxBB2 rule preserves the benefits observed in the quadratic case and the adaptive alternations of the BB1 and BoxBB2 selections still outperform the other rules. In the following, we describe preliminary results obtained on some box-constrained non-quadratic test problems, even if the general non-quadratic case deserves a specific analysis and it is beyond the scope of this work.

The test problems are generated as follows. Starting from an unconstrained minimization problem with a twice continuously differentiable objective function $g(x)$,

$$\min_{x \in \mathbb{R}^n} g(x), \tag{35}$$

for which a local minimum point x^* is known, the technique proposed in [35] is used to generate a box-constrained problem of the form

$$\min_{\ell \leq x \leq u} f(x) = g(x) + \sum_{i \in L} h_i(x_i) - \sum_{i \in U} h_i(x_i) \tag{36}$$

where $L = \{i \,|\, x_i^* = \ell_i\}$, $U = \{i \,|\, x_i^* = u_i\}$ and $h_i \colon \mathbb{R} \to \mathbb{R}$, $i \in L \cup U$, are twice continuously differentiable non-decreasing functions. Due to the special definition of $f(x)$, the point x^* is
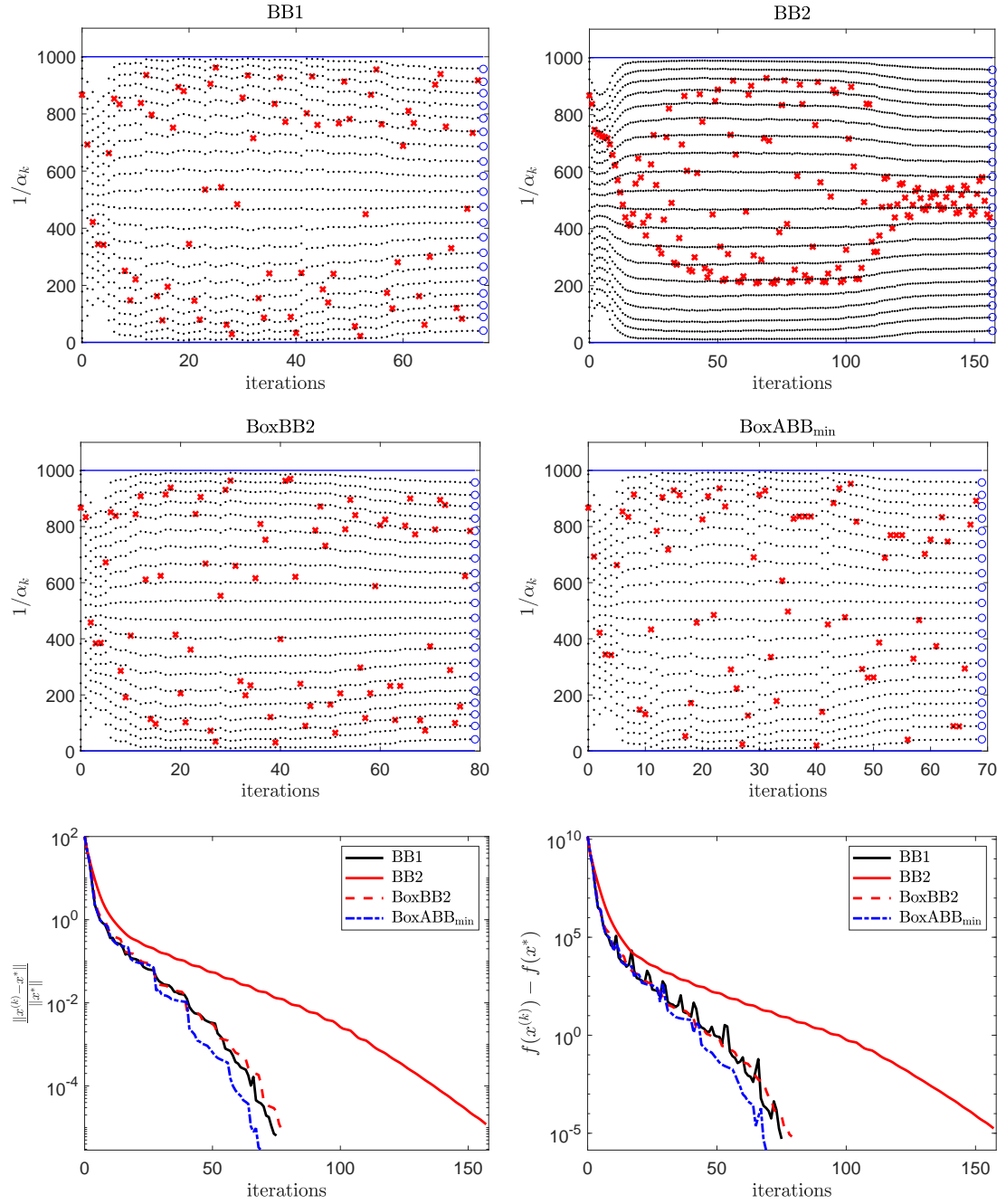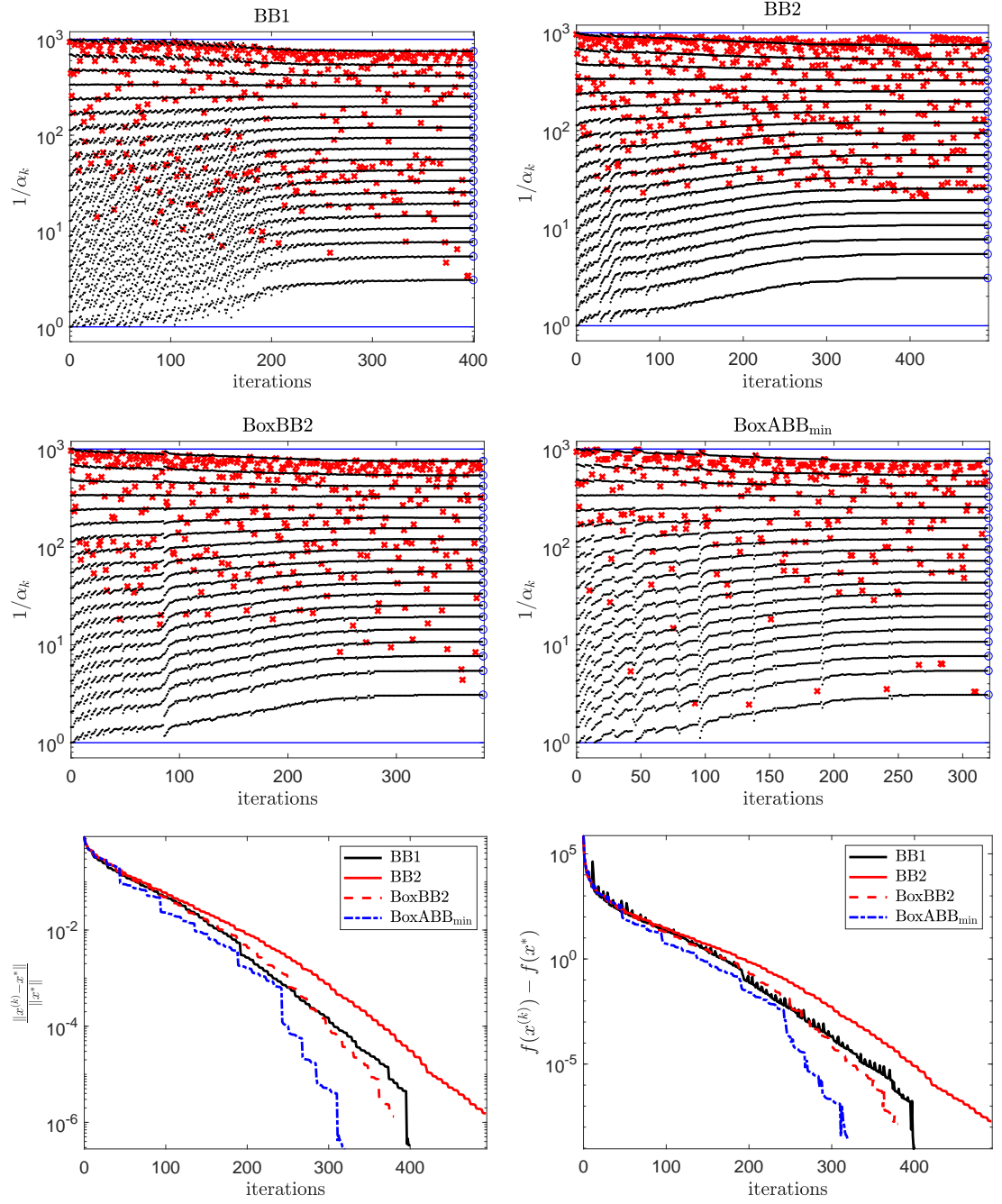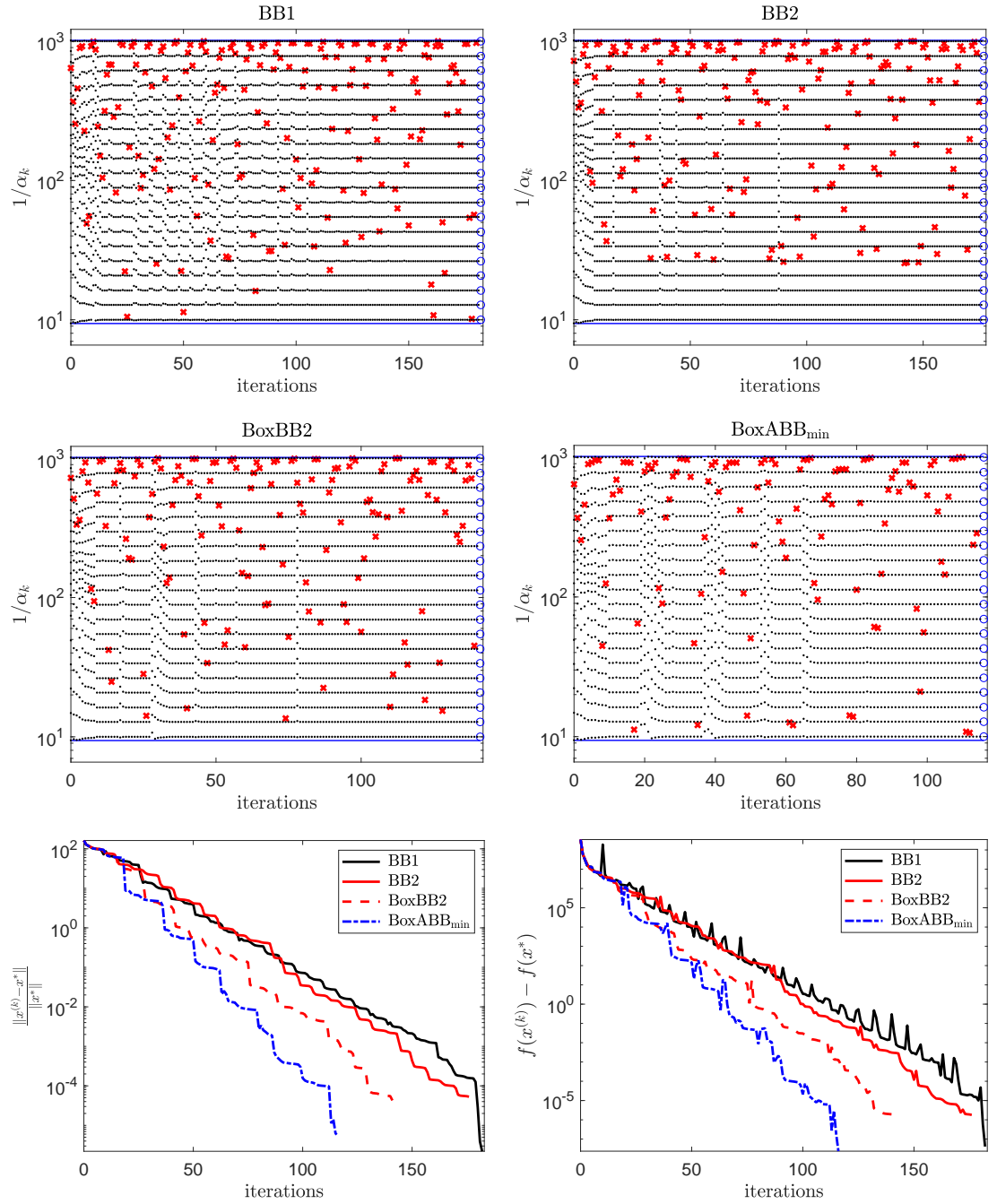
Figure 1: Behaviour of GP equipped with different steplength rules on TP1. Distribution of $\frac{1}{\alpha_k}$ with respect to the iterations for the BB1, BB2 (top panels), BoxBB2 and BoxABB$_{\text{min}}$ (middle panels) rules; errors on $x^{(k)}$ and $f(x^{(k)})$ for the different rules (bottom panels).
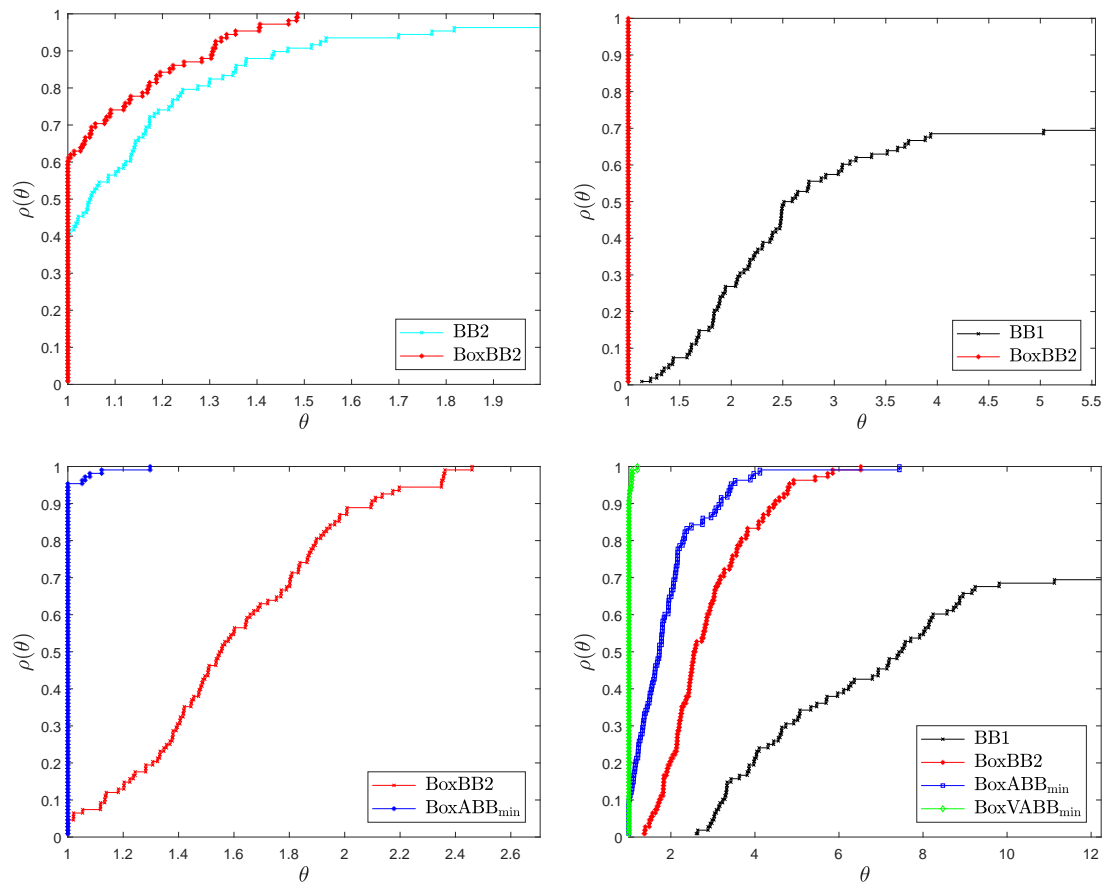
13

Figure 2: Behaviour of GP equipped with different steplength rules on TP2. Distribution of $\frac{1}{\alpha_k}$ with respect to the iterations for the BB1, BB2 (top panels), BoxBB2 and BoxABB$_{\min}$ (middle panels) rules; errors on $x^{(k)}$ and $f(x^{(k)})$ for the different rules (bottom panels).

Figure 3: Behaviour of GP equipped with different steplength rules on TP3. Distribution of $\frac{1}{\alpha_k}$ with respect to the iterations for the BB1, BB2 (top panels), BoxBB2 and BoxABB$_{\min}$ (middle panels) rules; errors on $x^{(k)}$ and $f(x^{(k)})$ for the different rules (bottom panels).
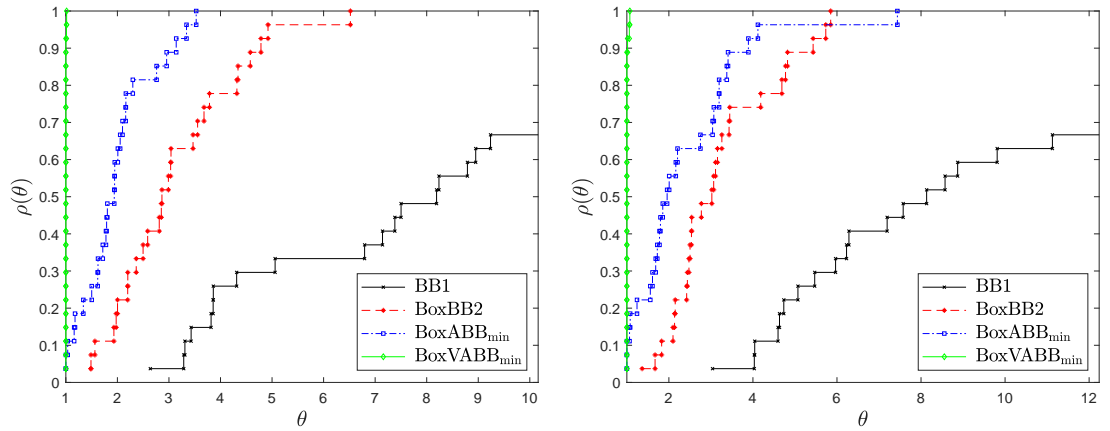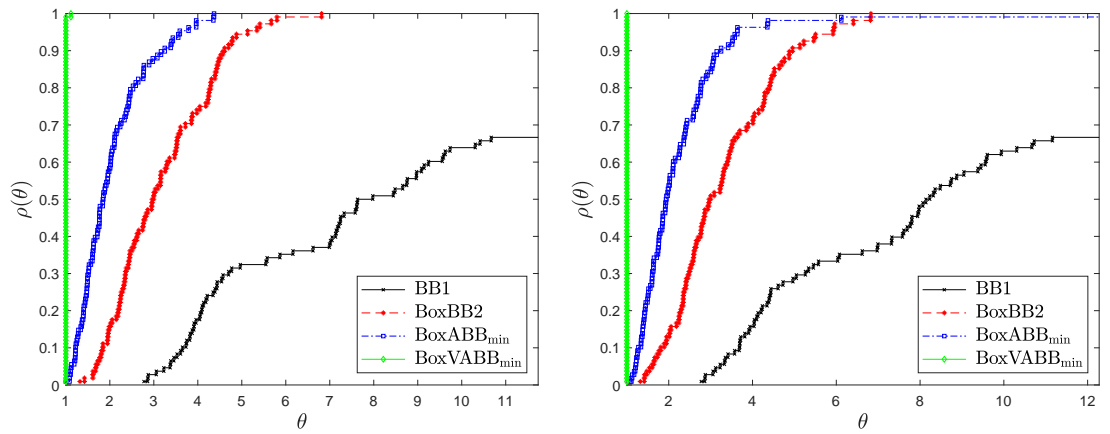
Figure 4: Runtime performance profiles obtained by the GP method equipped with different steplength rules on a set of box-constrained QP test problems with nondegenerate solutions.

Figure 5: Runtime performance profiles obtained by the GP method equipped with different steplength rules on the subsets of box-constrained QP test problems generated by setting $ndeg = 7$ (left panel) and $ndeg = 10$ (right panel).

Figure 6: Runtime performance profiles obtained by the GP method equipped with different steplength rules on a set of box-constrained QP test problems with degenerate solutions. Fraction of active variables at the solution that are degenerate equal to: $0.2 \cdot na$ (left panel) and $0.5 \cdot na$ (right panel).

a solution of (36). For our tests we selected some well-known non-quadratic functions $g(x)$, described below.

(i) Trigonometric function [36]:

$$g(x) = \|b - (Av(x) + Bu(x))\|^2,$$

where $v(x) = (\sin(x_1), ..., \sin(x_n))^T$, $u(x) = (\cos(x_1), ..., \cos(x_n))^T$, 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 $g(x^*) = 0$. The starting vector is set as $x^{(0)} = \min(u, (\max(\ell, x^* + 0.3\,r)))$, where $r \in \mathbb{R}^n$ has random entries from a uniform distribution in $[-\pi, \pi]$.

(ii) Chained Rosenbrock function [37]:

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

where $n = 500$, the values φ_i, $i = 1, \ldots, 50$, are defined as in [37, Table 1] and $\varphi_{i+50j} = \varphi_i$, $i = 1, ..., 50$, $j = 1, \ldots, 9$. In this case, a solution of the problem (35) is $x^* = (1, 1, ..., 1)^T$. The starting vector is set as $x^{(0)} = \min(u, (\max(\ell, x^* + 0.8\,r)))$, where $r \in \mathbb{R}^n$ has random entries from a uniform distribution in $[-1, 1]$.

(iii) Laplace2 function [10]:

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

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), $k, r, s = 1, \ldots, N$. 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) $d = 50$, $d_1 = 0.4$, $d_2 = 0.7$, $d_3 = 0.5$.

In both cases, the starting vector is $x^{(0)} = \frac{\ell+u}{2}$.

For each function $g(x)$, we built the corresponding constrained versions with the following choices for the functions $h_i(x)$, as suggested in [35]:

(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 $\beta_i = 10^{-\eta_i ndeg}$, where η_i is a random number in $(0, 1)$ and $ndeg = 1, 4, 10$; we recall that β_i are the Lagrangian multipliers associated to the active constraints and then the value $ndeg$ allows to control the degeneracy of the problem at x^*. The vectors ℓ and u are set in such a way that the number of active constraints at the solution is equal to a prefixed value na; the same number of lower and upper active constraints is assumed and different problems are generated by considering $na \approx 0.1 \cdot n, 0.5 \cdot n, 0.9 \cdot n$. The procedure just described led us to construct a total amount of 108 box-constrained non-quadratic test problems. Similarly to the previous experiments, Figure 7 shows the performance profiles obtained by solving the above set of non-quadratic problems by the GP method with different steplength rules. All the parameters involved by the GP scheme and by the steplength rules are set as in the quadratic case, except for the initial steplength, that now is $\alpha_0 = 1$, and the maximum number of iteration, that is equal to 4000. From Figure 7 we observe that the BoxBB2 rule confirms, also on these non-quadratic test problems, its better behaviour in comparison to the standard BB2 selection and well compares with the BB1 rule. Furthermore, the steplength strategies based on adaptive alternation of BB1 and BoxBB2 still provide the best performance. Then, by recalling that the proposed selections are designed essentially for achieving an effective approximation of the spectrum of the reduced Hessian, we may conclude that the spectral properties exhibited by the steplength rules play a crucial role for improving the gradient methods also in case of general box-constrained non-quadratic optimization problems.

4. Conclusions

In this work we developed a spectral analysis of the Barzilai-Borwein steplength rules in gradient projection methods for box-constrained QP problems and, as a consequence, we proposed a steplength strategy suitable to exploit information on the active constraints at each iteration. In particular, the new rule provides steplengths able to capture spectral properties of the reduced Hessian matrices, i.e., the submatrices of the Hessian of the objective function given by the intersection of the rows and the columns with indices corresponding to the inactive constraints at the iterations of a gradient projection scheme. This steplength selection seems very useful in achieving a fast annihilation of the gradient components that must be zeroed at the solution and, consequently, it can be efficiently exploited in combination with the standard Barzilai-Borwein rule within state of the art adaptive alternation steplength strategies. Preliminary numerical results suggest that the proposed rule shows benefits also in gradient projection methods for general box-constrained non-quadratic optimization problems. Future work will concern the generalization of this analysis to the cases of more general constraints or more complex schemes like the scaled gradient projection methods [26, 38, 22].
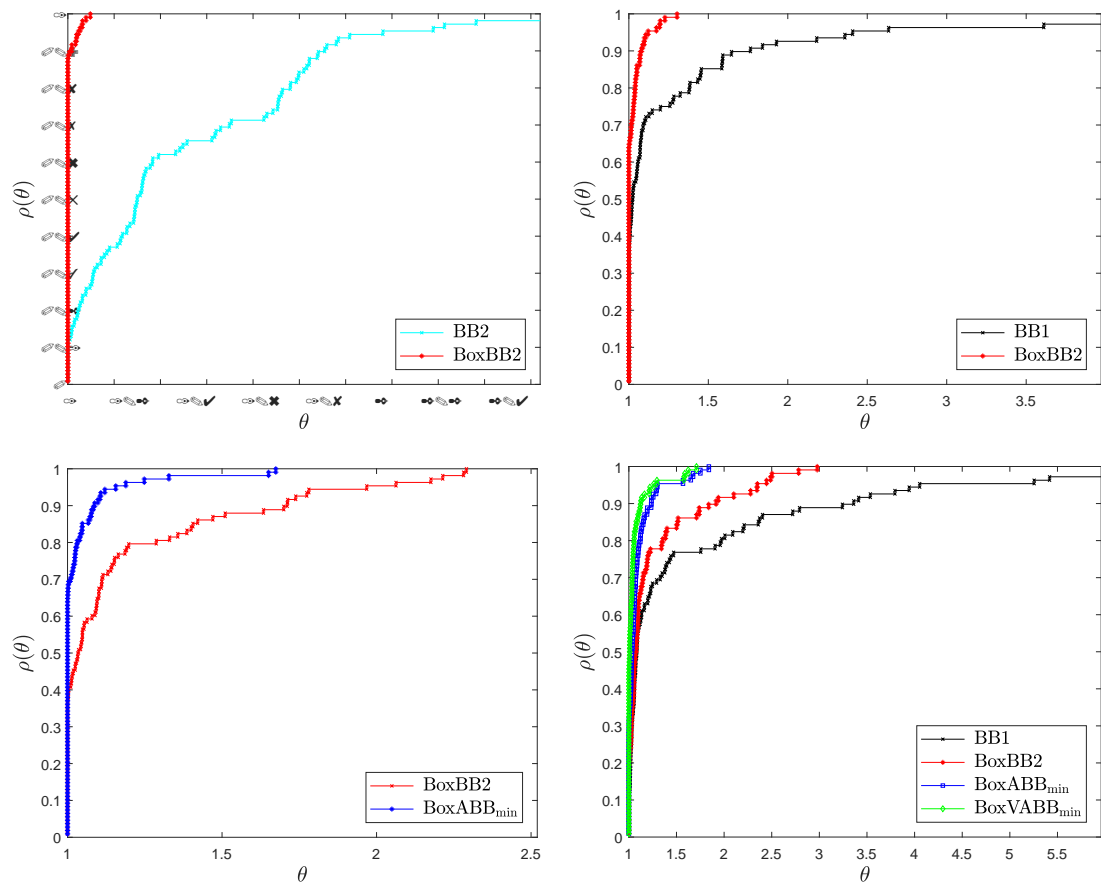
Figure 7: Runtime performance profiles obtained by the GP method equipped with different steplength rules on a set of box-constrained non-quadratic test problems.

References

[1] J. Barzilai, J. M. Borwein, Two-point step size gradient methods, IMA J. Numer. Anal. 8 (1988) 141–148.

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

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

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

[5] R. Fletcher, A limited memory steepest descent method, Math. Program., Ser. A 135 (2012) 413–436.

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

[7] M. Raydan, The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem, SIAM J. Optim. 7 (1) (1997) 26–33.

[8] Y. H. Dai, On the nonmonotone line search, J. Optim. Theory Appl. 112 (2002) 315–330.

[9] Y. H. Dai, L. Z. Liao, R-linear convergence of the Barzilai and Borwein gradient method, IMA J. Numer. Anal. 22 (1) (2002) 1–10.

[10] R. Fletcher, On the Barzilai-Borwein method, in: L. Qi, K. Teo, X. Yang, P. M. Pardalos, D. Hearn (Eds.), Optimization and Control with Applications, Vol. 96 of Applied Optimization, Springer, US, 2005, pp. 235–256.

[11] Y. H. Dai, W. W. Hager, K. Schittkowski, H. Zhang, The cyclic Barzilai-Borwein method for unconstrained optimization, IMA J. Num. Anal. 26 (3) (2006) 604–627.

[12] E. G. Birgin, J. M. Martínez, M. Raydan, Nonmonotone spectral projected gradient methods on convex sets, SIAM J. Optim. 10 (2000) 1196–1211.

[13] E. G. Birgin, J. M. Martínez, M. Raydan, Inexact spectral projected gradient methods on convex sets, IMA J. Numer. Anal. 23 (2003) 539–559.

[14] Y. H. Dai, R. Fletcher, Projected Barzilai-Borwein methods for large-scale box-constrained quadratic programming, Numer. Math. 100 (2005) 21–47.

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

[16] M. A. Figueiredo, R. D. Nowak, S. J. Wright, Projection for sparse reconstruction: Application to compressed sensing and other inverse problems, IEEE Journal of Selected Topics in Signal Processing 1 (2007) 586–597.

[17] M. Zhu, S. J. Wright, T. F. Chan, Duality-based algorithms for total-variation-regularized image restoration, Comput. Optim. Appl. 47 (3) (2008) 377–400.

[18] G. Yu, L. Qi, Y. H. Dai, On nonmonotone Chambolle gradient projection algorithms for total variation image restoration, J. Math. Imaging Vis. 35 (2009) 143–154.

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

[20] E. G. Birgin, J. M. Martínez, M. Raydan, Spectral projected gradient methods: Review and perspectives, J. Stat. Soft. 60 (3) (2014) 1–21.

[21] L. Antonelli, V. De Simone, D. di Serafino, On the application of the spectral projected gradient method in image segmentation, J. Math. Imaging Vis. 54 (1) (2015) 106–116.

[22] L. Pospíšil, Z. Dostál, The projected Barzilai-Borwein method with fall-back for strictly convex QCQP problems with separable constraints, Math. Comp. Simul. 145 (2018) 79–89.

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

[24] Y. Huang, H. Liu, On the rate of convergence of projected Barzilai-Borwein methods, Optim. Methods Softw. 30 (4) (2015) 880–892.

[25] A. Iusem, On the convergence properties of the projected gradient method for convex optimization, Comput. Appl. Math. 22 (1) (2003) 37–52.

[26] S. Bonettini, R. Zanella, L. Zanni, A scaled gradient projection method for constrained image deblurring, Inverse Probl. 25 (1) (2009) 015002.

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

[28] R. De Asmundis, D. di Serafino, F. Riccio, G. Toraldo, On spectral properties of steepest descent methods, IMA J. Numer. Anal. 33 (2013) 1416–1435.

[29] C. Gonzaga, R. M. Schneider, On the steepest descent algorithm for quadratic functions, Comput. Optim. Appl. 63 (2) (2016) 523–542.

[30] M. Prato, R. Cavicchioli, L. Zanni, P. Boccacci, M. Bertero, Efficient deconvolution methods for astronomical imaging: algorithms and IDL-GPU codes, Astron. Astrophys. 539 (2012) A133.

[31] R. Zanella, G. Zanghirati, R. Cavicchioli, L. Zanni, P. Boccacci, M. Bertero, G. Vicidomini, Towards real-time image deconvolution: application to confocal and sted microscopy, Sci. Rep. 3 (2013) 2523.

[32] J. J. Morè, G. Toraldo, Algorithms for bound constrained quadratic programming problems, Numer. Math. 55 (4) (1989) 377–400.

[33] D. di Serafino, G. Toraldo, M. Viola, J. L. Barlow, A two-phase gradient method for quadratic programming problems with a single linear constraint and bounds on the variables, SIAM Journal on Optimization. 28 (4) (2018) 2809–2838.

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

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

[36] R. Fletcher, M. J. D. Powell, A rapidly convergent descent method for minimization, Comput. J. 6 (1963) 163–168.

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

[38] F. Porta, M. Prato, L. Zanni, A new steplength selection for scaled gradient methods with application to image deblurring, J. Sci. Comp. 65 (2015) 895–919.