Steplength selections in gradient projection methods with applications to image deblurring

Serena Crisci

Università di Modena e Reggio Emilia - Dip. di Scienze Fisiche, Informatiche e Matematiche

serena.crisci@unimore.it

Abstract

In the framework of gradient projection-based approaches, steplength selection techniques are very crucial for the effectiveness of the method. In the context of constrained optimization, we propose modified versions of the well-known Barzilai-Borwein rules (and their extensions), highlighting their feature of capturing second-order information in a low cost way, as in the unconstrained case ([3]). The practical effectiveness of the proposed strategies has been tested on random large scale box-constrained quadratic problems, on some well-known non quadratic problems, and on image deblurring applications.

Mathematical Framework

Let consider the following box-constrained quadratic programming (BQP) problem

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

where $A \in \mathbb{R}^{n\times n}$ is symmetric positive definite, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$.
Let denote $g(x) \equiv \nabla f(x) = Ax - b$. We solve problem (1) by means of the gradient projection (GP) algorithm combined with a line-search strategy along the feasible direction. The main step of the GP method are described in Algorithm 1.

Initialization: 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 $k = 0, 1, \dots$ **do**
  $d^{(k)} = P_{\ell \le x \le u}\left(x^{(k)} - \alpha_k g(x^{(k)})\right) - x^{(k)}$; `// gradient projection step`
  $\lambda_k = 1; \quad f_{ref} = \max\{f(x^{k-i}), 0 \le i \le \min(k, M)\}$;
  while $f(x^{(k)} + \lambda_k d^{(k)}) > f_{ref} + \sigma \lambda_k g(x^{(k)})^T d^{(k)}$ **do**
    $\lambda_k = \delta \lambda_k$; `// backtracking step`
  end
  $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$;
  define the steplength $\alpha_{k+1} \in [\alpha_{min}, \alpha_{max}]$; `// steplength updating rule`
end

Algorithm 1: GP method for box-constrained quadratic programs

Steplength selection strategies

Unconstrained case. The standard Barzilai-Borwein (BB) rules [1] are obtained by imposing

$$\alpha_k^{BB1} = \arg\min_{\alpha} \left\| \alpha^{-1} s^{(k-1)} - y^{(k-1)} \right\| \quad \text{or} \quad \alpha_k^{BB2} = \arg\min_{\alpha} \left\| s^{(k-1)} - \alpha y^{(k-1)} \right\| \tag{2}$$

where $s^{(k-1)} = x^{(k)} - x^{(k-1)}$ and $y^{(k-1)} = g(x^{(k)}) - g(x^{(k-1)})$. From (2) we have:

$$\alpha_k^{BB1} = \frac{\|s^{k-1}\|^2}{(s^{k-1})^T y^{k-1}}, \quad \alpha_k^{BB2} = \frac{(s^{k-1})^T y^{k-1}}{\|y^{k-1}\|^2}. \tag{3}$$

Some well-known improvements of the BB rules are the strategies Alternate Barzilai-Borwein (ABB) [2] and its modification $\mathrm{ABB}_{\min}$ [4].

Box-constrained case: modified steplengths rules.

Let be $\mathcal{J} = \left\{ i \mid (x_i^{(k-1)} = \ell_i \wedge g_i^{(k-1)} \ge 0) \vee (x_i^{(k-1)} = u_i \wedge g_i^{(k-1)} \le 0) \right\}$ and $\mathcal{I} = \{1, ..., n\} - \mathcal{J}$; the problem related to BB1 rule can be formulated as

$$\min_{\alpha} \left\| \alpha^{-1} s_{\mathcal{J}}^{(k-1)} - y_{\mathcal{J}}^{(k-1)} \right\|^2 + \left\| \alpha^{-1} s_{\mathcal{I}}^{(k-1)} - y_{\mathcal{I}}^{(k-1)} \right\|^2.$$

Since $s_{\mathcal{J}}^{(k-1)} = 0$, only the term $\left\| \alpha^{-1} s_{\mathcal{I}}^{(k-1)} - y_{\mathcal{I}}^{(k-1)} \right\|^2$ affects the BB1 rule, so the effective computed value is:

$$\boxed{\alpha_k^{BB1} = \frac{\|s_{\mathcal{I}}^{k-1}\|^2}{(s_{\mathcal{I}}^{k-1})^T y_{\mathcal{I}}^{k-1}}} \tag{4}$$

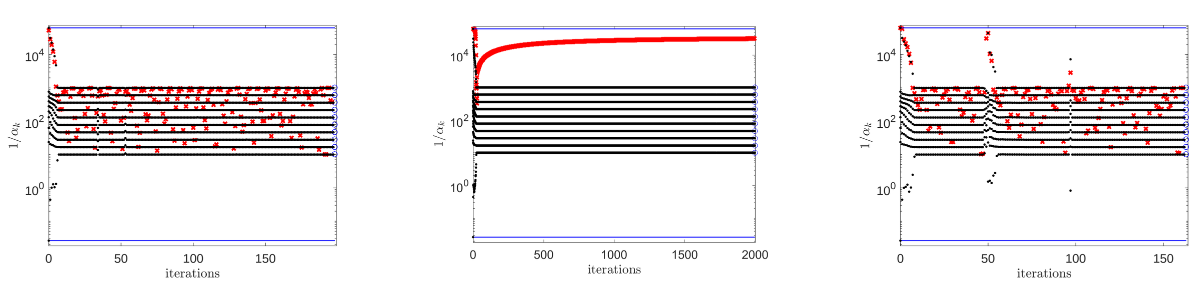
A similar argument on the BB2 steplength leads to the following formula

$$\alpha_k^{BB2} = \frac{(s_{\mathcal{I}}^{k-1})^T y_{\mathcal{I}}^{k-1}}{\|y_{\mathcal{I}}^{k-1}\|^2 + \|y_{\mathcal{J}}^{k-1}\|^2}. \tag{5}$$

Let denote by $A_{\mathcal{I},\mathcal{I}}$ the submatrix of A defined by the rows and columns with indices in $\mathcal{I}$, which we call *reduced Hessian matrix* at the $(k-1)$-th iteration. As proved in Theorem 1, $1/\alpha_k^{BB1}$ (4) belongs to the spectrum of $A_{\mathcal{I},\mathcal{I}}$, whereas $1/\alpha_k^{BB2}$ (5) might be outside of the spectrum of the reduced Hessian at $x^{(k-1)}$. We propose to correct the computed BB2 value as follows:

$$\boxed{\alpha_k^{MBB2} = \frac{(s_{\mathcal{I}}^{(k-1)})^T y_{\mathcal{I}}^{(k-1)}}{\|y_{\mathcal{I}}^{(k-1)}\|^2}} \tag{6}$$

Figure 1: BQP test problem of size $n = 1000$. Behaviour of $\frac{1}{\alpha_k}$ with respect to the iterations of GP equipped with BB1 rule (left), BB2 rule (central) and MBB2 rule (right); here a nonmonotone line search is used.



The next theorem states that steplengths (4)-(6) are the reciprocal of the Rayleigh quotients of $A_{\mathcal{I},\mathcal{I}}$.

Theorem 1

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

$$\lambda_{min}(A_{\mathcal{I},\mathcal{I}}) \le 1/\alpha_k^{BB1} \le \lambda_{max}(A_{\mathcal{I},\mathcal{I}}).$$
$$\lambda_{min}(A_{\mathcal{I},\mathcal{I}}) \le 1/\alpha_k^{MBB2} \le \lambda_{max}(A_{\mathcal{I},\mathcal{I}}).$$

As a consequence of the previous results, we can consider a modified $\mathrm{ABB}_{\min}$ scheme consisting in the alternation between BB1 and MBB2:

$$\boxed{\alpha_k^{MABB_{min}} = \begin{cases} \min\left(\alpha_j^{MBB2}, j = \max(1, k - m_a, ..., k)\right), & \text{if } \frac{\alpha_k^{MBB2}}{\alpha_k^{BB1}} < \tau \\ \alpha_k^{BB1} & \text{otherwise} \end{cases}} \tag{7}$$

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

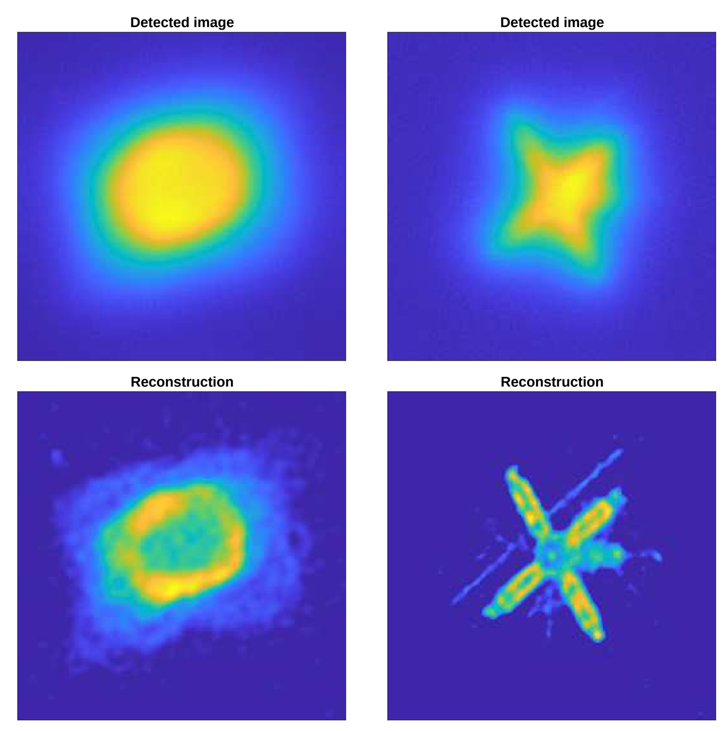
Application to image deblurring

Data affected by Gaussian noise

The test problems are generated by convolving the original 256×256 images (named **Nebula**, and **Spacecraft**) with a point spread function (simulated ground-based telescope http://www.mathcs.emory.edu/ nagy/RestoreTools/index.html), and perturbing the results with additive white Gaussian noise with variance 1 and zero background radiation [5]. The corresponding constrained minimization problem is a least squares problem with non negative constraints of the form:

$$\min_{x \ge 0} \frac{1}{2}\|Ax - y\|$$

where $y \in \mathbb{R}^{n^2}$ is the non-negative observed data, $A \in \mathbb{R}^{n^2 \times n^2}$ is the imaging matrix, and $x \in \mathbb{R}^{n^2}$ is the image to recover. Hereafter, we denote by RRE the relative reconstruction error, and by Ad-$\mathrm{ABB}_{\min}$ and Ad-$\mathrm{MABB}_{\min}$, respectively, the alternate approaches where the threshold τ is variable instead of being a constant parameter.



Method	Nebula			Spacecraft		
	It.	Time (s)	RRE	It.	Time (s)	RRE
ISRA	1903	72.28	0.074	2500	91.78	0.310
BB1	145	6.76	0.080	635	21.84	0.276
BB2	203	8.30	0.080	1048	32.86	0.276
MBB2	194	7.98	0.080	685	22.76	0.276
Ad-$\mathrm{ABB}_{\min}$	142	6.03	0.080	1853	56.25	0.276
Ad-$\mathrm{MABB}_{\min}$	137	4.48	0.080	1262	37.48	0.276

Table 1: First row panels: noisy and blurred images of Nebula (left), and Spacecraft (right). Second row panels: Nebula image recovered by GP method equipped with Ad-$\mathrm{MABB}_{\min}$ rule corresponding to the minimum RRE (left), Spacecraft image recovered by GP method equipped with MBB2 rule corresponding to the minimum RRE (right). Table: minimum RRE achieved by each algorithm, with the correspondings required number of iterations and execution time.

Data affected by Poisson noise

We considered three images of different size: a confocal microscopy phantom (**Micro**) [6], a spacecraft image (**Spacecraft**), and a microscopy phantom (**Tubule**) representing a micro-tubule network inside the cell, considered in [5]. The blurred and noisy images are obtained by convolving the original images with the point spread function described before, and by perturbing the result of the convolution with Poisson noise. Due to its features, the problem can be formulated as the minimization of a Kullback-Leibler divergence with a regularization term consisting in a smooth approximation of the total variation:

$$\min_{x \ge 0} \sum_{i=1}^{n^2} \left\{ y_i \log \frac{y_i}{(Ax+b)_i} + (Ax+b)_i - y_i \right\} + \beta \sum_{i,j=1}^{n} \sqrt{((\mathcal{D}x)_{i,j})_1^2 + ((\mathcal{D}x)_{i,j})_2^2 + \delta^2}$$

where $b \in \mathbb{R}^{n^2}$ is a known background radiation, $(Ax+b)_i > 0 \ \forall i = 1, \dots, n^2$, $\beta, \delta > 0$, and $\mathcal{D} \colon \mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$ is a discrete gradient operator, set through the standard finite difference scheme with periodic boundary conditions.

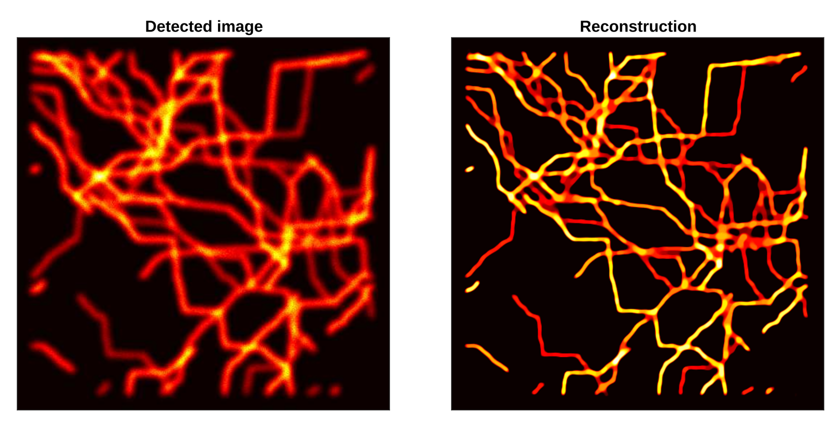


Image	n	Rule	It.	Time (s)	RRE
Micro	128	$\mathrm{ABB}_{\min}$	410	1.74	0.092
		$\mathrm{MABB}_{\min}$	370	1.54	0.091
Spacecraft	256	$\mathrm{ABB}_{\min}$	917	12.74	0.379
		$\mathrm{MABB}_{\min}$	666	9.56	0.375
Tubule	512	$\mathrm{ABB}_{\min}$	1472	246.10	0.575
		$\mathrm{MABB}_{\min}$	876	151.29	0.575

Table 2: Left: noisy and blurred Tubule image. Center: Tubule image recovered by GP method equipped with $\mathrm{MABB}_{\min}$ rule. Right: table reporting the RRE achieved by each algorithm, with the correspondings required number of iterations and execution time.

Forthcoming Research

- ✓ Generalization of the strategy to other feasible regions
- ✓ Steplengths selection rule based on Ritz-like values for constrained problems
- ✓ Analysis of the behaviour of modified steplength rules in presence of a variable metric

References

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

[2] Y. H. Dai and Y. Yuan. Alternate minimization gradient method. *IMA J. Numer. Anal.*, 23:377–393, 2003.

[3] 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.

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

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

[6] R. M. Willett and R. D. Nowak. Platelets: a multiscale approach for recovering edges and surfaces in photon limited medical imaging. *IEEE Trans. Med. Imaging*, 22:332–50, 2003.