Spectral Projected Gradient Methods

E. G. Birgin [*] J. M. Martínez [†] M. Raydan [‡]

January 17, 2007

Keywords: Spectral Projected Gradient Methods, projected gradients, nonmonotone line search, large scale problems, convex constrained problems, bound constrained problems, spectral gradient method.

MSC2000: 49M07, 49M10, 65K, 90C06, 90C20.

1 Introduction

Cauchy's steepest descent algorithm [22] is the most ancient method for multidimensional unconstrained minimization. Given f, a real smooth function defined on $I\!R^n$, the idea is to iterate according to:

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k), \tag{1}$$

with the expectancy that the sequence $\{x_k\}$ would approximate a minimizer of f. The greedy choice of the steplength α_k is

$$f(x_k - \alpha_k \nabla f(x_k)) = \min_{\alpha \geq 0} f(x_k - \alpha \nabla f(x_k)). \tag{2}$$

The poor practical behavior of (1)-(2) has been known for many years. If the level sets of f resemble long valleys, the sequence $\{x_k\}$ displays a typical zig-zagging trajectory and the speed of convergence is very slow. In the simplest case, in which f is a strictly convex quadratic, the method converges to the solution with a Q-linear rate of convergence whose factor tends to 1 when the condition number of the Hessian tends to infinity.

Nevertheless, the structure of the iteration (1) is very attractive, especially when one deals with large-scale (many variables) problems. Each iteration only needs the computation of the

[*]Department of Computer Science IME-USP, University of São Paulo, Rua do Matão 1010, Cidade Universitária, 05508-090, São Paulo SP, Brazil. This author was supported by PRONEX-Optimization (PRONEX - CNPq / FAPERJ E-26 / 171.164/2003 - APQ1), FAPESP (Grants 06/53768-0 and 03/09169-6) and CNPq (PROSUL 490333/2004-4). e-mail: egbirgin@ime.usp.br

[†]Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil. This author was supported by PRONEX-Optimization (PRONEX - CNPq / FAPERJ E-26 / 171.164/2003 - APQ1), FAPESP (Grant 06/53768-0) and CNPq. e-mail: martinez@ime.unicamp.br

[‡]Departamento de Computación, Facultad de Ciencias, Universidad Central de Venezuela, Ap. 47002, Caracas 1041-A, Venezuela (mraydan@reacciun.ve). Sponsored by the Center of Scientific Computing at UCV.

gradient $\nabla f(x_k)$ and the number of algebraic operations is linear in terms of n. As a consequence, a simple paper by Barzilai and Borwein published in 1988 [4] attracted some justified attention. Barzilai and Borwein discovered that, for some choices of α_k, Cauchy's method converges superlinearly to the solution, if $f : I\!R^2 \to I\!R$ is a convex quadratic. Some members of the optimization community began to believe that the existence of an efficient method for large-scale minimization based only on gradient directions could be possible.

In 1993, Raydan [60] proved the convergence of the Barzilai-Borwein method for arbitrary strictly convex quadratics. He showed that the method was far more efficient than the steepest descent algorithm (1)-(2) although it was not competitive with the Conjugate Gradient method of Hestenes and Stiefel [49] for quadratic problems. The possibility of obtaining superlinear convergence for arbitrary n was discarded by Fletcher's work [40] (see also [60]) and a bizarre behavior of the method seemed to discourage the application to general (not necessarily quadratic) unconstrained minimization: the sequence of functional values $f(x_k)$ did not decrease monotonically and, sometimes, monotonicity was severely violated.

However, starting with the work by Grippo, Lampariello and Lucidi [47], nonmonotone strategies for function minimization began to become popular. These strategies made it possible to define globally convergent algorithms without monotone decrease requirements. The philosophy behind nonmonotone strategies is that, many times, the first choice of a trial point by a minimization algorithm hides a lot of wisdom about the problem structure and that such knowledge can be destroyed by the decrease imposition. For example, if one applies Newton's method to a problem in which several components of the gradient are linear, these components vanish at the first trial point of each iteration, but the objective function value does not necessarily decrease at this trial point.

Therefore, the conditions were given for the implementation of the Barzilai-Borwein method for general unconstrained minimization with the help of a nonmonotone strategy. Raydan [61] defined this method in 1997 using the GLL strategy [47]. He proved global convergence and exhibited numerical experiments that showed that, perhaps surprisingly, the method was more efficient than classical conjugate gradient methods for minimizing general functions. These nice comparative numerical results were possible because, albeit the Conjugate Gradient method of Hestenes and Stiefel continued to be the rule of choice for solving many convex quadratic problems, its efficiency was hardly inherited by generalizations for minimizing general functions. Therefore, there existed a wide space for variations of the Barzilai-Borwein idea.

The Spectral Projected Gradient (SPG) method [16, 17, 18] was born from the marriage of the Barzila-Borwein (spectral) nonmonotone ideas with classical projected gradient strategies [7, 46, 53]. This method is applicable to convex constrained problems in which the projection on the feasible set is easy to compute. Since its appearance, the method has been intensively used in applications [3, 6, 10, 14, 15, 19, 20, 24, 26, 35, 42, 50, 59, 63, 64, 65, 69]. Moreover, it has been the object of several spectral-parameter modifications, alternative nonmonotone strategies have been suggested, convergence and stability properties have been elucidated and it has been combined with other algorithms for different optimization problems.

2 Method

2.1 The secant connection

Quasi-Newton secant methods for unconstrained optimization [36, 37] obey the recursive formula

$$x_{k+1} = x_k + \alpha_k B_k^{-1} \nabla f(x_k). \tag{3}$$

The sequence of matrices $\{B_k\}$ satisfy the *secant equation*

$$B_{k+1} s_k = y_k, \tag{4}$$

where $s_k = x_{k+1} - x_k$ and $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$. By (4), it can be shown that, at the *trial point* $x_k - B_k^{-1}\nabla f(x_k)$, the affine approximation of $\nabla f(x)$ that interpolates the gradient at x_k and x_{k-1} vanishes for all $k \geq 1$.

Now assume that we want a matrix B_{k+1} with a very simple structure that satisfies (4). More precisely, we wish

$$B_{k+1} = \sigma_{k+1} I,$$

with $\sigma_{k+1} \in I\!R$. The equation (4) becomes:

$$\sigma_{k+1} s_k = y_k.$$

In general, this equation cannot be solved. However, accepting the least-squares solution that minimizes $\|\sigma s_k - y_k\|_2^2$, we obtain:

$$\sigma_{k+1} = \frac{s_k^T y_k}{s_k^T s_k}. \tag{5}$$

This formula defines the most popular Barzilai-Borwein method [61]. Namely, the method for unconstrained minimization is of the form (3), where, at each iteration,

$$d_k = -\frac{1}{\sigma_k} \nabla f(x_k)$$

and formula (5) is used to generate the coefficients σ_k provided that they are bounded away from zero and that they are not very large. In other words, the method uses safeguards $0 < \sigma_{\min} < \sigma_{\max} < \infty$ and defines, at each iteration:

$$\sigma_{k+1} = \min\{\sigma_{\max}, \max\{\sigma_{\min}, \frac{s_k^T y_k}{s_k^T s_k}\}\}.$$

By the Mean-Value Theorem of Integral Calculus, one has:

$$y_k = \left(\int_0^1 \nabla^2 f(x_k + t s_k) dt\right) s_k.$$

Therefore, formula (5) defines a Rayleigh quotient relative to the average Hessian matrix $(\int_0^1 \nabla^2 f(x_k + t s_k) dt) s_k$. This coefficient is between the minimum and the maximum eigenvalue of the average Hessian, which motivates the denomination of Spectral Method [16].

Writing the secant equation as $H y_k = s_k$, which is also standard in the Quasi-Newton tradition, we arrive to a different spectral coefficient: $\frac{y_k^T y_k}{s_k^T y_k}$. Curiously, both this dual and the primal coefficient had been used for many years in practical quasi-Newton methods to define the initial matrices B_0 [58].

2.2 The line search

The Spectral Projected Gradient method (SPG) aims to minimize f on a closed and convex set Ω. The method, as well as its unconstrained counterpart [61] has the form

$$x_{k+1} = x_k + \alpha_k d_k. \tag{6}$$

The search direction d_k has been defined in [16] as

$$d_k = P(x_k - \frac{1}{\sigma_k}\nabla f(x_k)) - x_k,$$

where P denotes the Euclidean projection on Ω. A related method with approximate projections has been defined in [18]. The direction d_k is a descent direction, which means that, if $d_k \neq 0$, one has that $f(x_k + \alpha d_k) \ll f(x_k)$ for α small enough. This means that, in principle, one could define convergent methods imposing sufficient decrease at every iteration. However, this leads to disastrous practical results. For this reason, the spectral methods employ a nonmonotone line search that does not impose functional decrease at every iteration. In [16, 17, 18] the GLL [47] search is used. This line search depends on an integer parameter $M \geq 1$ and imposes a functional decrease every M iterations (if $M = 1$ then GLL line search reduces to a monotone line search).

The line search is based on a safeguarded quadratic interpolation and aims to satisfy an Armijo-type criterion with a sufficient decrease parameter $\gamma \in (0,1)$. The safeguarding procedure acts when the minimum of the one-dimensional quadratic lies outside $[\tau_1, \tau_2\alpha]$, and not when it lies outside $[\tau_1\alpha, \tau_2\alpha]$ as usually implemented. This means that, when interpolation tends to reject 90% (for $\sigma_1 = 0.1$) of the original search interval ($[0,1]$), we judge that its prediction is not reliable and we prefer the more conservative bisection. The complete line search procedure is described below.

Algorithm 3.1: Line search

Compute $f_{\max} = \max\{f(x_{k-j}) \mid 0 \leq j \leq \min\{k, M-1\}\}$, $x_+ \leftarrow x_k + d_k$, $\delta \leftarrow \langle \nabla f(x_k)), d_k\rangle$ and set $\alpha \leftarrow 1$.

Step 1. *Test nonmonotone Armijo-like criterion*

If $f(x_+) \leq f_{\max} + \gamma\alpha\delta$ then set $\alpha_k \leftarrow \alpha$ and stop.

Step 2. *Compute a safeguarded new trial steplength*

Compute $\alpha_{\text{tmp}} \leftarrow -\frac{1}{2}\alpha^2\delta/(f(x_+) - f(x_k) - \alpha\delta)$.

If $\alpha_{\text{tmp}} \geq \sigma_1$ and $\alpha_{\text{tmp}} \leq \sigma_2\alpha$ then set $\alpha \leftarrow \alpha_{\text{tmp}}$. Otherwise, set $\alpha \leftarrow \alpha/2$.

Step 3. Compute $x_+ \leftarrow x_k + \alpha d_k$ and go to Step 1.

Remark. In the case of rejection of the first trial point, the next ones are computed along the same direction. As a consequence, the projection operation is performed only once per iteration.

2.3 General form and convergence

The Spectral Projected Gradient method SPG aims to solve the problem

$$\text{Minimize } f(x) \text{ subject to } \quad x \in \Omega, \tag{7}$$

where f admits continuous first derivatives and $\Omega \subset I\!R^n$ is closed and convex.

We say that a point $x \in \Omega$ is *stationary* if

$$\nabla f(x)^T d \geq 0$$

for all $d \in I\!R^n$ such that $x + d \in \Omega$.

In [18], SPG method has been presented as a member of a wider family of "Inexact Variable Metric" methods for solving (7). Let $I\!B$ be the set of $n \times n$ positive definite matrices such that $\|B\| \leq L$ and $\|B^{-1}\| \leq L$. Therefore, $I\!B$ is a compact set of $I\!R^{n\times n}$. In the spectral gradient approach, the matrices will be of the form σI, with $\sigma \in [\sigma_{\min}, \sigma_{\max}]$.

Algorithm 4.1: Inexact Variable Metric Method

Assume $\eta \in (0, 1]$, $\gamma \in (0, 1)$, $0 < \tau_1 < \tau_2 < 1$, $M \geq 1$ an integer number. Let $x_0 \in \Omega$ be an arbitrary initial point. We denote $g_k = \nabla f(x_k)$ for all $k = 0, 1, 2, \ldots$. Given $x_k \in \Omega$, $B_k \in I\!B$, the steps of the $k-$th iteration of the algorithm are:

Step 1. *Compute the search direction*

Consider the subproblem

$$\text{Minimize } Q_k(d) \quad \text{subject to} \quad x_k + d \in \Omega, \tag{8}$$

where

$$Q_k(d) = \frac{1}{2} d^T B_k d + g_k^T d.$$

Let $\bar{d}_k$ be the minimizer of (8). (This minimizer exists and is unique by the strict convexity of the subproblem (8), but does not need to be computed.)

Let d_k be such that $x_k + d_k \in \Omega$ and

$$Q_k(d_k) \leq \eta\, Q_k(\bar{d}_k).$$

If $d_k = 0$, stop the execution of the algorithm declaring that x_k is a stationary point.

Step 2. *Compute the steplength*

Compute $f_{\max} = \max\{f(x_{k-j}) \mid 0 \leq j \leq \min\{k, M-1\}\}$, $\delta \leftarrow \langle g_k, d_k \rangle$ and set $\alpha \leftarrow 1$.

If

$$f(x_k + \alpha d_k) \le f_{\max} + \gamma\alpha\delta, \tag{9}$$

set $\alpha_k = \alpha$, $x_{k+1} = x_k + \alpha_k d_k$ and finish the iteration. Otherwise, choose $\alpha_{\text{new}} \in [\tau_1\alpha, \tau_2\alpha]$, set $\alpha \leftarrow \alpha_{\text{new}}$ and repeat test (9).

Remarks. (a) Algorithm 3.1 is a particular case of Step 2 of Algorithm 4.1. (b) In the definition of Algorithm 4.1 the possibility $\eta = 1$ corresponds to the case in which the subproblem (8) is solved exactly.

The main theoretical results [18] are stated below. Firstly, it is shown that an iteration necessarily finishes and then it is shown that every limit point of a sequence generated by the algorithm is stationary.

Theorem 4.1. *The algorithm is well defined.*

Theorem 4.2. *Assume that the level set $\{x \in \Omega \mid f(x) \le f(x_0)\}$ is bounded. Then, either the algorithm stops at some stationary point x_k, or every limit point of the generated sequence is stationary.*

2.4 Numerical example

In [17] a family of *location* problems was introduced. Given a set of *npol* disjoint polygons in $I\!R^2$ we wish to find the point y that minimizes the sum of the distances to those polygons. Therefore, the problem is

$$\min_{z^i, y} \sum_{i=1}^{npol} \|z^i - y\|_2$$

$$\text{subject to} \quad z^i \in P_i,\ i = 1, \ldots, npol.$$

Let us write $x = (z_1^1, z_2^1, \ldots, z_1^{npol}, z_2^{npol}, y_1, y_2)$. We observe that the problem has $2 \times (npol + 1)$ variables. The number of (linear inequality) constraints is $\sum_{i=1}^{npol} \nu_i$, where ν_i is the number of vertices of the polygon P_i. Each constraint defines a half-plane in $I\!R^2$. Figure 1 shows the solution of a small five-polygons problem.

For projecting x onto the feasible set observe that we only need to project each z^i separately onto the corresponding polygon P_i. In the projection subroutine we consider the half-planes that define the polygon. If z^i belongs to all these half-planes, then z^i is the projection onto P_i. Otherwise, we consider the set of half-planes to which z^i does not belong. We project z^i onto these half-planes and we discard the projected points that do not belong to P_i. Let A_i be the (finite) set of nondiscarded half-plane projections and let V_i be the set of vertices of P_i. Then, the projection of z^i onto P_i is the point of $A_i \cup V_i$ that is closest to z^i. The projection subroutine are included in the test driver for SPG method [17].

Varying *npol* and choosing randomly the localization of the polygons and the number of vertices of each one, several test problems were generated and solved by the SPG method in [17]. The biggest problem had 48,126 polygons, 96,254 variables and 578,648 constraints. Using

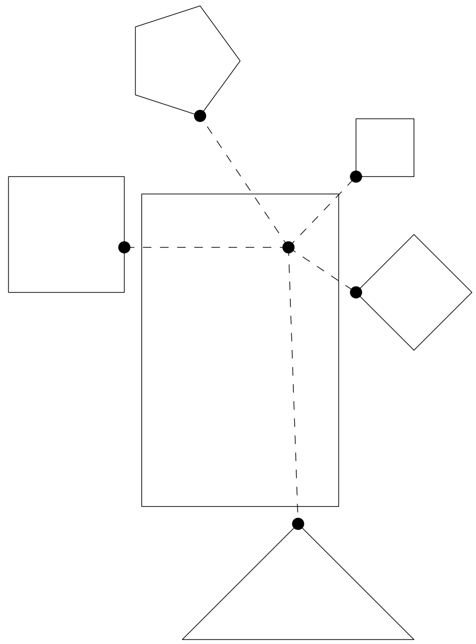

Figure 1: Five-polygons problem.

the origin as initial approximation, it was solved by the SPG method in 17 iterations, using 19 function evaluations and 12.97 seconds of CPU time on a Sun SparcStation 20 with the following main characteristics: 128Mbytes of RAM, 70MHz, 204.7 mips, 44.4 Mflops. (Codes were in Fortran 77 and the compiler option adopted was "-O".)

2.5 Further developments

Developments on spectral gradient and spectral projected gradient ideas include:

1. Application and implementation of the spectral methods to particular optimization problems: Linear inequality constraints were considered in [1]. In [38] the SPG method was used to solve Augmented Lagrangian subproblems. The spectral gradient method solves the subproblems originated by the application of an exponential penalty method to linearly constrained optimization in [56].

2. Preconditioning: The necessity of preconditioning for very ill-conditioned problems has

been recognized in several works [5, 23, 45, 54, 57].

3. Extensions: The spectral residual direction was defined in [51] to introduce a new method that aims to solve nonlinear systems of equations using only the vector of residues. See, also, [48, 52, 70]. The SPG method has been extended for solving non-differentiable convex constrained problems [25].

4. Association with other methods: The SPG method has been used in the context of active-set methods for box-constrained optimization in [2, 13, 12]. Namely, SPG iterations are used in these methods for abandoning faces whereas Newtonian iterations are employed inside the faces of the feasible region. The opposite idea was used in [44], where spectral directions were used inside the faces and a different orthogonal strategy was employed to modify the set of current active constraints (see also [9]). Spectral ideas were also used in association with conjugate gradients in [11]. Combinations of the spectral gradient method with other descent directions were suggested in [21, 28].

5. Nonmonotone alternative rules: Dai and Fletcher [30] observed that, in some cases, even the descent GLL strategy was very conservative and, so, more chance should be given to the pure spectral method ($\alpha_k = 1$ for all k). However, they showed that, for quadratic minimization with box constraints, the pure method is not convergent. Therefore, alternative tolerant nonmonotone rules were suggested. Dai and Zhang [31] noted that the behavior of spectral methods heavily depend on the choice of the parameter M of the GLL search and proposed an adaptive nonmonotone search independent of M. Over and under relaxations of the spectral step were studied by Raydan and Svaiter [62].

6. Alternative choices of the spectral parameter: In [43] it was observed that the convergence theory of the spectral gradient method for quadratics remains valid when one uses Rayleigh coefficients originated in retarded iterations (see also [55]). In [32], for unconstrained optimization problems, the same stepsize is reused for m consecutive iterations (CBB method). This cyclic method is locally linearly convergent to a local minimizer with positive definite Hessian. Numerical evidence indicates that when $m > n/2 >= 3$, where n is the problem dimension, CBB is locally superlinearly convergent. In the special case $m = 3, n = 2$, the convergence rate is, in general, no better than linear [32].

   In [34] the stepsize in the spectral gradient method was interpreted from the point of view of interpolation and two competitive modified spectral-like gradient methods were defined. Yuan [68] defined a new stepsize for unconstrained optimization that seems to possess spectral gradient properties preserving monotonicity.

7. Asymptotic behavior analysis: A careful consideration of the asymptotic practical and theoretical behavior of the Barzilai-Borwein method may be found in [41]. Dai and Fletcher [29] showed interesting transition properties of the spectral gradient method for quadratic functions as depending on the number of variables. Dai and Liao [33] proved the R-linear convergence of the spectral gradient method for general functions and, as a consequence, established that the spectral stepsize is always accepted by the non-monotone line search when the iterate is close to the solution. The convergence of the inexact SPG method was

established in [66, 67] under different assumptions than the ones used in [18]. Assuming Lipschitz-continuity of the objective functions, these authors eliminated the bounded-level-set assumption of [18].

References

[1] R. Andreani, E. G. Birgin, J. M. Martínez and J. Yuan, Spectral projected gradient and variable metric methods for optimization with linear inequalities, *IMA Journal of Numerical Analysis* 25, pp. 221–252, 2005.

[2] M. Andretta, E. G. Birgin and J. M. Martínez, Practical active-set Euclidian trust-region method with spectral projected gradients for bound-constrained minimization, *Optimization* 54, pp. 305–325, 2005.

[3] D. Azofeifa, N. Clark and W. Vargas, Optical and electrical properties of terbium films as a function of hydrogen concentration, *Physica Status Solidi B – Basic Solid State Physics* 242, pp. 2005–2009, 2005.

[4] J. Barzilai and J. M. Borwein, Two point step size gradient methods, *IMA Journal of Numerical Analysis* 8, pp. 141–148, 1988.

[5] L. Bello and M. Raydan, Preconditioned spectral projected gradient method on convex sets, *Journal of Computational Mathematics* 23, pp. 225–232, 2005.

[6] L. Bello and M. Raydan, Convex constrained optimization for the seismic reflection tomography problem, *Journal of Applied Geophysics* 62, pp. 158–166, 2007.

[7] D. P. Bertsekas, On the Goldstein-Levitin-Polyak gradient projection method, *IEEE Transactions on Automatic Control* 21, pp. 174–184, 1976.

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

[9] R. H. Bielschowsky, A. Friedlander, F. A. M. Gomes, J. M. Martínez and M. Raydan, An adaptive algorithm for bound constrained quadratic minimization, *Investigación Operativa* 7, pp. 67–102, 1997.

[10] E. G. Birgin, I. E. Chambouleyron and J. M. Martínez, Optimization problems in the estimation of parameters of thin films and the elimination of the influence of the substrate, *Journal of Computational and Applied Mathematics* 152, pp. 35–50, 2003.

[11] E.G. Birgin and J. M. Martínez, A spectral conjugate gradient method for unconstrained optimization, *Applied Mathematics and optimization* 43, pp. 117–128, 2001.

[12] E. G. Birgin and J. M. Martínez, Large-scale active-set box-constrained optimization method with spectral projected gradients, *Computational Optimization and Applications* 23, pp. 101–125, 2002.

[13] E. G. Birgin and J. M. Martínez, A box constrained optimization algorithm with negative curvature directions and spectral projected gradients, *Computing [Supl]* 15, pp. 49–60, 2001.

[14] E. G. Birgin, J. M. Martínez, W. F. Mascarenhas and D. P. Ronconi, Method of sentinels for packing items within arbitrary convex regions, *Journal of the Operational Research Society* 57, pp. 735–746, 2006.

[15] E. G. Birgin, J. M. Martínez, F. H. Nishihara, D. P. Ronconi, Orthogonal packing of rectangular items within arbitrary convex regions by nonlinear optimization, *Computers and Operations Research* 33, pp. 3535–3548, 2006.

[16] E. G. Birgin, J. M. Martínez and M. Raydan, Nonmonotone spectral projected gradient methods on convex sets, *SIAM Journal on Optimization* 10, pp. 1196–1211, 2000.

[17] E. G. Birgin, J. M. Martínez and M. Raydan, Algorithm 813: SPG - Software for convex-constrained optimization, *ACM Transactions on Mathematical Software* 27, pp. 340–349, 2001.

[18] E. G. Birgin, J. M. Martínez and M. Raydan. Inexact Spectral Projected Gradient methods on convex sets, *IMA Journal on Numerical Analysis* 23, pp. 539–559, 2003.

[19] E. G. Birgin, J. M. Martínez and D. P. Ronconi, Minimization subproblems and heuristics for an applied clustering problem, *European Journal of Operational Research* 146, pp. 19–34, 2003.

[20] E. G. Birgin, J. M. Martínez and D. P. Ronconi, Optimizing the packing of cylinders into a rectangular container: A nonlinear approach, *European Journal of Operational Research* 160, pp. 19–33, 2005.

[21] C. Brezinski, Block descent methods and hybrid procedures for linear systems, *Numerical Algorithms* 29, pp. 21–32, 2002.

[22] A. Cauchy, Méthode générale pour la résolution des systèmes d'équations simultanées, *Compte-Rendu de l'Académie des Sciences* 27, pp. 536–538, 1847.

[23] J. P. Chehab and M. Raydan, Implicit and adaptive inverse preconditioned gradient methods for nonlinear problems, *Applied Numerical Mathematics* 55, pp. 32–47, 2005.

[24] D. Cores and M. Loreto, A generalized two-point ellipsoidal anisotropic ray tracing for converted waves, *Optimization and Engineering*, to appear.

[25] A. Crema, M. Loreto and M. Raydan, Spectral projected subgradient with a momentum term for the Lagrangean dual approach, *Computers & Operations Research* 34, pp. 3174–3186, 2007.

[26] F. Curiel, W. E. Vargas and R. G. Barrera, Visible spectral dependence of the scattering and absorption coefficients of pigmented coatings from inversion of diffuse reflectance spectra, *Applied Optics* 41, pp. 5969–5978, 2002.

[27] Y. H. Dai [2000], On the nonmonotone line search, *Journal of Optimization Theory and Applications* 112, pp. 315–330, 2002.

[28] Y. H. Dai, Alternate step gradient method, *Optimization* 52, pp. 395–415, 2003.

[29] Y. H. Dai and R. Fletcher, On the asymptotic behaviour of some new gradient methods, *Mathematical Programming* 103, pp. 541–559, 2005.

[30] Y. H. Dai and R. Fletcher, Projected Barzilai-Borwein methods for large-scale box-constrained quadratic programming, *Numerische Mathematik* 100, pp. 21–47, 2005.

[31] Y. H. Dai and H. C. Zhang, Adaptive two-point stepsize gradient algorithm, *Numerical Algorithms* 27, pp. 377–385, 2001.

[32] Y.H. Dai, W. W. Hager, K. Schittkowski and H. C. Zhang, The cyclic Barzilai-Borwein method for unconstrained optimization, *IMA Journal of Numerical Analysis* 26, pp. 604–627, 2006.

[33] Y. H. Dai and L. Z. Liao, R-linear convergence of the Barzilai and Borwein gradient method, *IMA Journal on Numerical Analysis* 22, pp. 1–10, 2002.

[34] Y. H. Dai, J. Y. Yuan and Y. X. Yuan, Modified two-point stepsize gradient methods for unconstrained optimization, *Computational Optimization and Applications* 22, pp. 103–109, 2002.

[35] G.P. Deidda, E. Bonomi and C. Manzi, Inversion of electrical conductivity data with Tikhonov regularization approach: some considerations, *Annals of Geophysics* 46, pp. 549–558, 2003.

[36] J. E. Dennis and J. J. Moré, Quasi-Newton methods, motivation and theory, *SIAM Review* 19, pp. 46–89, 1977.

[37] J. E. Dennis and R. B. Schnabel, *Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, NJ, 1983.

[38] M. A. Diniz-Ehrhardt, M. A. Gomes-Ruggiero, J. M. Martínez and S. A. Santos, Augmented Lagrangian algorithms based on the spectral projected gradient method for solving nonlinear programming problems, *Journal of Optimization Theory and Applications* 123, pp. 497–517, 2004.

[39] R. Fletcher, *Practical methods of Optimization*, John Wiley, 1987.

[40] R. Fletcher, Low storage methods for unconstrained optimization, *Lectures in Applied Mathematics (AMS)*, 26, pp. 165–179, 1990.

[41] R. Fletcher, On the Barzilai-Borwein method, Department of Mathematics, University of Dundee, NA/207, Dundee, Scotland, 2001.

[42] J. B. Francisco, J. M. Martínez and L. Martínez, Density-based globally convergent trust-region methods for self-consistent field electronic structure calculations, *Journal of Mathematical Chemistry* 40, pp. 349–377, 2006.

[43] A. Friedlander, J. M. Martínez, B. Molina and M. Raydan, Gradient method with retards and generalizations, *SIAM Journal on Numerical Analysis* 36, pp. 275–289, 1998.

[44] A. Friedlander, J. M. Martínez and M. Raydan, A new method for large-scale box constrained convex quadratic minimization problems, *Optimization Methods and Software* 5, pp. 57–74, 1995.

[45] W. Glunt, T. L. Hayden and M. Raydan, Preconditioners for distance matrix algorithms, *J. Comp. Chem.*, 15, pp. 227–232, 1994.

[46] A. A. Goldstein, Convex Programming in Hilbert Space, *Bulletin of the American Mathematical Society* 70, pp. 709–710, 1964.

[47] L. Grippo, F. Lampariello and S. Lucidi, A nonmonotone line search technique for Newton's method, *SIAM Journal on Numerical Analysis* 23, pp. 707–716, 1986.

[48] L. Grippo and M. Sciandrone, Nonmonotone Derivative Free Methods for Nonlinear Equations, *Computational Optimization and Applications*, to appear (DOI: 10.1007/s10589-007-9028-x).

[49] M. R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, *J. Res. NBS* 49, pp. 409–436, 1952.

[50] Z. Jiang, Applications of conditional nonlinear optimal perturbation to the study of the stability and sensitivity of the Jovian atmosphere, *Advances in Atmospheric Sciences* 23, pp. 775–783, 2006.

[51] W. La Cruz and M. Raydan, Nonmonotone spectral methods for large-scale nonlinear systems, *Optimization Methods and Software* 18, pp. 583–599, 2003.

[52] W. La Cruz, J. M. Martínez and M. Raydan, Spectral residual method without gradient information for solving large-scale nonlinear systems of equations, *Mathematics of Computations* 75, pp. 1429–1448, 2006.

[53] E. S. Levitin and B. T. Polyak, Constrained Minimization Problems, *USSR Computational Mathematics and Mathematical Physics* 6, pp. 1–50, 1966.

[54] F. Luengo, M. Raydan, W. Glunt and T. L. Hayden, Preconditioned Spectral Gradient Method, *Numerical Algorithms* 30, pp. 241–258, 2002.

[55] F. Luengo and M. Raydan, Gradient method with dynamical retards for large-scale optimization problems, *Electronic Transactions on Numerical Analysis (ETNA)* 16, pp. 186–193, 2003.

[56] J. M. Martínez, E. A. Pilotta and M. Raydan, Spectral gradient methods for linearly constrained optimization, *Journal of Optimization theory and Application* 125, pp. 629–651, 2005.

[57] B. Molina and M. Raydan, Preconditioned Barzilai-Borwein method for the numerical solution of partial differential equations, *Numerical Algorithms* 13, pp. 45–60, 1996.

[58] S. S. Oren, On the selection of parameters in self-scaling variable metric algorithms, *Mathematical Programming* 7, pp. 351–367, 1974.

[59] A. Ramirez-Porras and W. E. Vargas-Castro, Transmission of visible light through oxidized copper films: feasibility of using a spectral projected gradient method, *Applied Optics* 43, pp. 1508–1514, 2004.

[60] M. Raydan, On the Barzilai and Borwein choice of steplength for the gradient method, *IMA Journal of Numerical Analysis* 13, pp. 321–326, 1993.

[61] M. Raydan, The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem, *SIAM Journal on Optimization* 7, pp. 26–33, 1997.

[62] M. Raydan and B. F. Svaiter, Relaxed steepest descent and Cauchy-Barzilai-Borwein method, *Computational Optimization and Applications* 21, pp. 155–167, 2002.

[63] T. Serafini, G. Zanghirati and L. Zanni, Gradient projection methods for quadratic programs and applications in training support vector machines, *Optimization Methods and Software* 20, pp. 347–372, 2005.

[64] W. E. Vargas, Inversion methods from Kiabelka-Munk analysis, *Journal of Optics A - Pure and Applied Optics* 4, pp. 452–456, 2002.

[65] W. E. Vargas, D. E. Azofeifa and N. Clark, Retrieved optical properties of thin films on absorbing substrates from transmittance measurements by application of a spectral projected gradient method, *Thin Solid Films* 425, pp. 1–8, 2003.

[66] C. Y. Wang and Q. Liu, Convergence properties of inexact projected gradient methods, *Optimization* 55, pp. 301–310, 2006.

[67] C. Y. Wang, Q. Liu and X. M. Yang, Convergence properties of nonmonotone spectral projected gradient methods, *Journal of Computational and Applied Mathematics* 182, pp. 51–66, 2005.

[68] Y-X. Yuan, A new stepsize for the steepest descent method, *Journal of Computational Mathematics* 24, pp. 149–156, 2006.

[69] N. Zeev, O. Savasta and D. Cores, Non-monotone spectral projected gradient method applied to full waveform inversion, *Geophysical Prospecting* 54, pp. 525–534, 2006.

[70] L. Zhang and W. J. Zhou, Spectral gradient projection method for solving nonlinear monotone equations, *Journal of Computational and Applied Mathematics* 196, pp. 478–484, 2006.

[71] B. Zhou, L. Gao and Y. H. Dai, Gradient methods with adaptive step-sizes, *Computational Optimization and Applications* 35, pp. 69–86, 2006.

[72] B. Zhou, L. Gao and Y. H. Dai, Monotone projected gradient methods for large-scale box-constrained quadratic programming, *Science in China Series A - Mathematics* 49, pp. 688–702, 2006.