Journal of Computational Mathematics, Vol.23, No.3, 2005, 225–232.

PRECONDITIONED SPECTRAL PROJECTED GRADIENT METHOD ON CONVEX SETS *1)

Lenys Bello
(*Dpto. de Matemática, Facultad de Ciencias (FACYT), Universidad de Carabobo, Valencia, Venezuela*)

Marcos Raydan
(*Dpto. de Computación, Universidad Central de Venezuela, Ap. 47002, Caracas 1041-A, Venezuela*)

Abstract

The spectral gradient method has proved to be effective for solving large-scale unconstrained optimization problems. It has been recently extended and combined with the projected gradient method for solving optimization problems on convex sets. This combination includes the use of nonmonotone line search techniques to preserve the fast local convergence. In this work we further extend the spectral choice of steplength to accept preconditioned directions when a good preconditioner is available. We present an algorithm that combines the spectral projected gradient method with preconditioning strategies to increase the local speed of convergence while keeping the global properties. We discuss implementation details for solving large-scale problems.

Mathematics subject classification: 49M, 90C, 65K.
Key words: Spectral gradient method, Projected gradient method, Preconditioning techniques, Nonmonotone line search.

1. Introduction

We consider the optimization problem

$$\text{minimize } \{f(x) : x \in \Omega\},$$

where Ω is a nonempty closed and convex set in $\Re^n$, n is large, f is continuously differentiable, $g(x) = \nabla f(x)$ is available, and $G(x) \approx \nabla^2 f(x)$ is also available and will be considered as a preconditioner. Our main objective is to develop a preconditioned and projected extension of the spectral gradient method to solve this problem.

Spectral gradient methods are nonmonotone schemes that have recently received considerable attention in the numerical analysis and optimization literature. They were introduced by Barzilai and Borwein [1], the convergence for quadratics was established by Raydan [17], and more recently, a proof of the R-linear rate of convergence for convex quadratics was given by Dai and Liao [10]. A complete review is presented by Fletcher [11], and the asymptotic behavior is studied by Dai and Fletcher [9].

The spectral gradient methods have been applied succesfully to find local minimizers of large scale problems ([4, 5, 7, 8, 12, 18, 20]), to solve box-constrained quadratic optimization (Bielschowsky et al. [2]), and also to minimize general smooth functions on convex sets [6]. Preconditioned spectral gradient versions have also been developed ([16, 13, 15]). However, the combination of preconditioning techniques to accelerate the process and projected techniques on convex sets, for robustness and regularity, has not been studied.

* Received December 10, 2003; final revised July 17, 2004.
1) This author was partially supported by UCV-PROJECT 97-003769.

The paper is divided into sections as follows. In Section 2 we describe briefly the preconditioned spectral gradient method, and we recall some of the important aspects of the spectral projected gradient method on convex sets. In Section 3 we present the new algorithm that combines both ideas taking into account the difficult cases. In Section 4 we present preliminary and encouraging numerical results, and some final remarks.

2. Previous Extensions

We describe briefly the most important properties of the preconditioned spectral gradient method and the spectral projected gradient method on convex sets.

2.1 Preconditioned spectral gradient method

The iterates of the Preconditioned Spectral Gradient (PSG) method presented by Glunt, Hayden, and Raydan [13] are defined by

$$x_{k+1} = x_k - \alpha_k^{-1} z_k \ ,$$

where $z_k = G_k^{-1} g_k$, G_k is a nonsingular approximation to the Hessian of f at x_k and the scalar α_k is given by

$$\alpha_k = (-\alpha_{k-1}) \frac{z_{k-1}^t y_{k-1}}{z_{k-1}^t g_{k-1}} \ ,$$

where x_0 and α_0 are given initial data (see also [16]).

The PSG method requires no line search during the process but does not guarantee monotonic descent in the objective function. As a consequence, Raydan [18] proposed a globalization scheme for the spectral gradient algorithm that fits nicely with the nonmonotone behavior of this family of methods. Roughly speaking, the algorithm forces the step to satisfy this weak condition:

$$f(x_{k+1}) \leq \max_{0 \leq j \leq M} f(x_{k-j}) + \gamma g_k^t (x_{k+1} - x_k) \ ,$$

where M is a nonnegative integer and γ is a small positive number. When $M > 0$ this condition allows the objective function to increase at some iterations and still guarantees global convergence. This globalization strategy is based on the nonmonotone line search technique of Grippo, Lampariello and Lucidi [14].

A direct combination of the PSG method and the nonmonotone globalization strategy described above produces an algorithm fully described in Luengo et. al. [15].

2.2 Spectral projected gradient method

There have been many different variations of the projected gradient method that can be viewed as the constrained extensions of the optimal gradient method for unconstrained minimization. They all have the common property of maintaining feasibility of the iterates by frequently projecting trial steps on the feasible convex set. In particular, Birgin et al. [6, 3] combine the projected gradient method with recently developed ingredients in optimization, as follows. The algorithm starts with $x_0 \in \Re^n$ and is based on the spectral projected gradient direction $d_k = P(x_k - \alpha_k g(x_k)) - x_k$, where α_k is the spectral choice of steplength $\frac{\langle s_{k-1}, s_{k-1} \rangle}{\langle s_{k-1}, y_{k-1} \rangle}$, and for $z \in \Re^n$, $P(z)$ is the projection on Ω. In the case of rejection of the first trial point, $x_k + d_k$, the next ones are computed along the same direction, i.e., $x_+ = x_k + \lambda d_k$, using a nonmonotone line search to force the following condition

$$f(x_+) \leq \max_{0 \leq j \leq \min\{k, M-1\}} f(x_{k-j}) + \gamma \lambda \langle d_k, g(x_k) \rangle,$$

where $M \geq 1$ is a given integer. As a consequence, the projection operation must be performed only once per iteration. More details can be found in [6] and [3].

3. New Algorithm

We combine both versions, preconditioned and projected, to obtain a new algorithm. However, the combination cannot be obtained by simply connecting both ideas. The most important difficulty is that the preconditioned direction, even if it is a descent direction, might not be a descent direction after being projected on Ω. Therefore, additional controls need to be included to produce a robust and well-defined algorithm. In the rest of this document the notation $\|z\|$ represents the Euclidean norm of the vector z.

PSPG: Preconditioned Spectral Projected Gradient Algorithm
Given $x_0 \in \Omega$, $M > 0$, $\alpha_0 \in [\epsilon, 1/\epsilon]$, $\gamma \in (0,1)$, $0 < \sigma_1 < \sigma_2 < 1$, a small tolerance $tol > 0$ for the stopping criterion and a tolerance *tolpre* for activating the preconditioner; $0 < \epsilon < 1$ and $0 < c < 1$.

Set $k \leftarrow 0$, $\hat{d}_0 = P(x_0 - \alpha_0 g(x_0)) - x_0$ and precond=off

while $(\|\hat{d}_k\| > tol)$

  if $(\|\hat{d}_k\| \leq tolpre)$ then (precond=on)
  end if

  if (precond=on) choose G_k and
    solve $G_k d_k = g_k$ for d_k,
    else $d_k = g_k$
  end if

  set $d_k = P(x_k - \alpha_k d_k) - x_k$,

  if (precond=on) and $(d_k^t g_k > -\epsilon \max(\|d_k\|\,\|\hat{d}_k\|, \|d_k\|^2, \|g_k\|^2))$, then
    precond=off, $tolpre = tolpre * c$ and $d_k = \hat{d}_k$
  end if

  set $\lambda \leftarrow 1$ and set $x_+ = x_k + d_k$
  while $(f(x_+) > \max_{0 \leq j \leq \min\{k, m-1\}}\{f(x_{k-j})\} + \gamma\lambda\langle d_k, g(x_k)\rangle)$
    choose $\sigma \in [\sigma_1, \sigma_2]$, set $\lambda = \sigma\lambda$ and set $x_+ = x_k + \lambda d_k$
  endwhile

  set $x_{k+1} \leftarrow x_+$, set $s_k = x_{k+1} - x_k$, set $\lambda_k = \lambda$,

  set $y_k = g(x_{k+1}) - g(x_k)$ and set $b_k = \langle d_k, y_k\rangle$.

  If $(b_k \leq \epsilon)$ then set $\alpha_{k+1} = \frac{1}{\epsilon}$
    else, set $a_k = s_k^t g_k$ and set $\alpha_{k+1} = \min\{\frac{1}{\epsilon}, \max\{\epsilon, \frac{a_k}{b_k}\}\}$.
  end if

  set $k \leftarrow k+1$ and $\hat{d}_k = P(x_k - \alpha_k g_k) - x_k$
endwhile

set $x_* \leftarrow x_k$.

Remarks.
1) Notice that if $\Omega = \Re^n$ then the PSPG algorithm reduces to the PSG algorithm in [6]; and if $G_k = I$ for all k then it reduces to the SPG method described in [15].
2) The preconditioner should be activated when the iterations are "close" to a solution ([15]). In general, depending on the specific application, this information is supplied by the user. In the PSPG algorithm we use a default criterion, i.e., $\|\hat{d}_k\| \leq tolpre$, where *tolpre* is given by the user. Notice that we do not impose any conditions on the matrix G_k.
3) In order to save computations involved by invoking the preconditioner too early and often,

we suggest that when the preconditioner is turned off by the above criterion, then the strictness of the local test be increased, for example by lowering the parameter *tolpre* by the factor c, as indicated in the algorithm.
4) If the search direction d_k does not satisfy the sufficient decrease condition

$$d_k^t g_k \leq -\epsilon \max(\|d_k\| \, \|\hat{d}_k\|, \|d_k\|^2, \|g_k\|^2), \tag{1}$$

then we use $\hat{d}_k$ as a search direction, that is also used for the stopping criterion, and so this choice does not represent an additional computational cost.

Theorem 3.1. *Assume that $\Omega_0 = \{x \in \Omega : f(x) \leq f(x_0)\}$ is a bounded set. Then Algorithm PSPG is well defined, and either $\hat{d}_j = 0$ for some finite j or any accumulation point of the sequence $\{x_k\}$ that it generates is a constrained stationary point.*

Proof. In Algorithm PSPG the search direction is either $d_k = P(x_k - \alpha_k G_k^{-1} g_k) - x_k$ or $\hat{d}_k = P(x_k - \alpha_k g_k) - x_k$. When the preconditioner is activated we try d_k, but if it does not satisfy the sufficient decrease condition (1), then we use $\hat{d}_k$. If the preconditioner is not activated we also use $\hat{d}_k$. Since we force $\alpha_k \in [\epsilon, 1/\epsilon]$ for all k, then by Lemma 2.1 in [6], when using $\hat{d}_k$ the sufficient decrease condition (1) is satisfied. Therefore, for all k the search direction satisfies (1), and the nonmonotone line search procedure terminates in a finite number of trials, which implies that the algorithm is well defined.

Now, we make use of the first part of the proof of the convergence theorem in [14, p. 709]. Let us define $m(k) = \min(k, M)$. Clearly, $m(0) = 0$ and

$$0 \leq m(k) \leq \min(m(k-1) + 1, M) \quad \text{for } k \geq 1 \, .$$

Using the fact that all search directions d_k satisfies (1),then we can obtain positive numbers c_1 and c_2 such that they satisfy $g_k^t d_k \leq -c_1 \|g_k\|^2$, and $\|d_k\| \leq c_2 \|g_k\|$ for all k. In fact, from (1) we have $d_k^t g_k \leq -\epsilon \|g_k\|^2$ for all k, and $c_1 = \epsilon$. Once again using (1) we obtain $d_k^t g_k \leq -\epsilon \|d_k\|^2$, and by the Cauchy-Schwarz inequality it follows that

$$\epsilon \, \|d_k\|^2 \leq |d_k^t g_k| \leq \|d_k\| \, \|g_k\|,$$

which implies that

$$\|d_k\| \leq \frac{1}{\epsilon} \|g_k\|,$$

and hence $c_2 = 1/\epsilon$. Finally, in our algorithm the trial steps are all constant ($a = 1$). Therefore, repeating the same arguments in [14, p. 710-711] we obtain:

$$\lim_{k \to \infty} \lambda_k (g_k^t d_k) = 0. \tag{2}$$

Let $\bar{x} \in \Omega$ be an accumulation point of $\{x_k\}$, relabel $\{x_k\}$ a subsequence converging to $\bar{x}$, and $\{\lambda_k\}$, $\{d_k\}$, and $\{\hat{d}_k\}$ the corresponding subsequences generated in the algorithm. We consider two cases:

Case 1. Assume that $\inf \lambda_k = 0$. Suppose, by contradiction, that $\bar{x}$ is not a constrained stationary point. Hence, by continuity and compactness, there exists $\delta > 0$ such that

$$\|P(\bar{x} - \alpha g(\bar{x})) - \bar{x}\| > \frac{\delta}{\epsilon} > 0 \quad \text{for all} \quad \alpha \in [\epsilon, \frac{1}{\epsilon}].$$

Since the condition (1) is satisfied by all iterates of algorithm PSPG, then

$$\langle g(x_k), \frac{d_k}{\|d_k\|} \rangle < -\epsilon \|\hat{d}_k\| < -\frac{\delta}{2} \quad \text{for all} \quad \alpha \in [\epsilon, \frac{1}{\epsilon}], \tag{3}$$

and k large enough on the subsequence that converges to $\bar{x}$.

Since $\inf \lambda_k = 0$, there exists a subsequence $\{x_k\}_K$ such that

$$\lim_{k\in K} \lambda_k = 0.$$

In that case, from the way λ_k is chosen in the nonmonotone line search, there exists an index $\bar{k}$ sufficiently large such that for all $k \geq \bar{k}$, $k \in K$, there exists ρ_k, $0 < \sigma_1 \leq \rho_k \leq \sigma_2$, for which $\lambda_k/\rho_k > 0$ fails to satisfy the nonmonotone line search condition, i.e.,

$$f(x_k + \frac{\lambda_k}{\rho_k} d_k) > \max_{0\leq j\leq M-1} f(x_{k-j}) + \gamma \frac{\lambda_k}{\rho_k} \langle g(x_k), d_k\rangle \geq f(x_k) + \gamma \frac{\lambda_k}{\rho_k} \langle g(x_k), d_k\rangle.$$

Hence,

$$\frac{f(x_k + \frac{\lambda_k}{\rho_k} d_k) - f(x_k)}{\lambda_k/\rho_k} > \gamma \langle g(x_k), d_k\rangle.$$

By the mean value theorem, this relation can be written as

$$\langle g(x_k + t_k d_k), d_k\rangle > \gamma \langle g(x_k), d_k\rangle, \quad \text{for all } k \in K,\ k \geq \bar{k}, \tag{4}$$

where t_k is a scalar in the interval $[0, \lambda_k/\rho_k]$ that goes to zero as $k \in K$ goes to infinity.

Taking a convenient subsequence such that $d_k/\|d_k\|$ is convergent to d, and taking limits in (4) we deduce that $(1-\gamma)\langle g(\bar{x}), d\rangle \geq 0$. (In fact, observe that $\{\|d_k\|\}_K$ is bounded and so $t_k\|d_k\| \to 0$.) Since $(1-\gamma) > 0$ and $\langle g(x_k), d_k\rangle < 0$ for all k, then $\langle g(\bar{x}), d\rangle = 0$.
By continuity and the definition of d_k this implies that for k large enough on that subsequence we have that

$$\langle g(x_k), \frac{d_k}{\|d_k\|}\rangle > -\delta/2,$$

which contradicts (3).

Case 2. Assume that $\inf \lambda_k \geq \rho > 0$. Hence, by (2) we obtain

$$\lim_{k\to\infty} g_k^t d_k = 0,$$

which implies, using (1), that $\lim_{k\to\infty} g_k = 0$. Therefore, $g(\bar{x}) = 0$, and

$$\|P(\bar{x} - \alpha g(\bar{x})) - \bar{x}\| = 0 \quad \text{for all} \quad \alpha \in [\epsilon, \frac{1}{\epsilon}],$$

which implies that $\bar{x}$ is a constrained stationary point.

4. Numerical Experiments

We compare the Spectral Projected Gradient Method (SPG) and the Preconditioned Spectral Projected Gradient Method (PSPG) on 8 standard test problems that can be found in the literature. A description of the functions and the starting points can be found in [18] and references therein. All the experiments were run on a PC Pentium III, 800 MHz and 256 Mbytes RAM, Fortran 77 (double precision), we used $\gamma = 10^{-4}$, $M = 10$, $\sigma_1 = 0.1$, $\sigma_2 = 0.6$, $\epsilon = 10^{-20}$, $\alpha_0 = 1/\|g_0\|$ and we stop when $\|\hat{d}_k\| \leq 10^{-6}$. For the PSPG we used the tridiagonal part of the Hessian as a preconditioner. In this case, solving the preconditioned linear system requires $O(n)$ flops and the storage for the matrix G_k is only 2 n-dimensional vectors. The preconditioner is activated ("local test") whenever $\|P_\Omega(x_k - \alpha_k g_k) - x_k\| \leq tolpre$, where $tolpre$ is a tolerance factor. If the preconditioning strategy is deactivated ($precond = off$), then we set $tolpre = tolpre * 10^{-1}$. In our particular experiments, Ω is a box, i.e.,

$$\Omega = \{x \,:\, l \leq x \leq u\}$$

where the vectors l and u are given and they can have infinite values in some of the positions. Table 1 lists the 8 functions and the structure of their Hessians. However, due to space restrictions, we only report three cases from this table. We report the best case, the worst case, and an average case that we observed in our experiments. In all the experiments we verified that the two methods converged to the same point. We report in tables 2,3, and 4 the numerical results for the three functions. In particular, we report the dimension of the problem (n), the tolerance factor for the "local test" ($tolpre$), the number of times that the preconditioner was activated (Precond(t)) where t is the iteration at which it was activated for the last time, the number of iterations required for convergence ($iter$), the number of function evaluations (F) the number of gradient evaluations (G), the vectors (l) and (u), the required CPU time in seconds (Time), and whether the unconstrained solution x^* is in Ω or not.

We observe that the global PSPG method is a robust method to find local minimizers of nonquadratic functions subject to bound constraints. It outperforms the SPG method in number of iterations and computational work.

For the Strictly Convex 2 function the preconditioning matrix is the exact Hessian (diagonal) and so the global PSPG method shows superlinear convergence in the last few iterations. This fact explains the excellent behavior of the PSPG for that particular test problem. For the Penalty 1 function, that is not convex, is highly nonlinear, and the exact Hessian is dense, the tridiagonal part of the Hessian is not a very accurate preconditioner. As a consequence, we observe the worst behavior of the PSPG method when compared with the SPG method on the 8 standard test problems. Even in that case, the PSPG method is competitive in number of iterations and computational work. Finally, for the Extended Powell Singular function we oberved what we have called the average behavior. In that case, the PSPG method converges in fewer number of iterations and less computational work than the SPG method. Indeed, in the average, when the solution is in Ω the PSPG method is approximately six time faster than the SPG method, and when the solution is not in Ω, it is approximately twice as fast as the SPG method.

In general, we observe that it is important to activate the preconditioner at the right iteration by using correctly the parameter $tolpre$, which plays an important role during the convergence process. Of course, in order to establish definite conclusions it is necessary to run experiments on real and larger problems, using more realistic preconditioners. These are all interesting topics that deserve further investigation in the near future.

In summary, our preliminary numerical results indicate that the PSPG method combines in a suitable way the preconditioning technique and the projected scheme to produce a promising idea that accelerates the convergence while adding robustness and regularity to the process, in the sense of [19].

Table 1: Standard test functions.

Function	Name	Hessian
1	Brown Almost Linear	dense
2	Broyden Tridiagonal	tridiagonal
3	Oren's Power	dense
4	Penalty 1	dense
5	Extended Powell Singular	pentadiagonal
6	Extended Rosenbrock	tridiagonal
7	Variably Dimensioned	dense
8	Strictly Convex 2	diagonal

Table 2: PSPG with and without preconditioning for Strictly Convex 2 ($x^* = 0$) for different values of n and different values of l and u.

n	*tolpre*	Precond(t)	*iter*	F	G	l	u	Time	$x^* \in \Omega$
100	1.0d-20	off	83	99	84	-10	10	0.23	yes
100	1.0d+10	1(1)	7	8	8	-10	10	0.01	yes
500	1.0d-20	off	214	286	215	$-\infty$	0.5	0.93	yes
500	1.0d+10	1(1)	6	7	7	$-\infty$	0.5	0.01	yes
1000	1.0d-20	off	366	549	367	$-\infty$	0.5	2.12	yes
1000	1.0d+10	1(1)	6	7	7	$-\infty$	0.5	0.02	yes
100	1.0d-20	off	78	82	79	-40	$10(u_1 = -3, u_n = 6)$	0.17	no
100	1.0d+10	1(1)	7	8	8	-40	$10(u_1 = -3, u_n = 6)$	0.02	no
1000	1.0d-20	off	347	475	348	-40	$10(u_1 = -3, u_n = 6)$	1.92	no
1000	1.0d+10	1(1)	7	8	8	-40	$10(u_1 = -3, u_n = 6)$	0.06	no
10000	1.0d-20	off	1466	2253	1467	-40	$10(u_1 = -3, u_n = 6)$	57.07	no
10000	1.0d+10	1(1)	7	8	8	-40	$10(u_1 = -3, u_n = 6)$	0.21	no

Table 3: PSPG with and without preconditioning for Penalty 1 ($x^* \approx 0.02$) for different values of n and different values of l and u.

n	*tolpre*	Precond(t)	*iter*	F	G	l	u	Time	$x^* \in \Omega$
100	1.0d-20	off	154	398	155	-10	10	0.33	yes
100	1.0d-2	4(65)	78	133	79	-10	10	0.19	yes
500	1.0d-20	off	50	69	51	-10	10	0.20	yes
500	1.0d-3	3(51)	56	86	57	-10	10	0.25	yes
100	1.0d-20	off	75	373	76	$-10(l_1 = 5)$	10	0.20	no
100	1.0d-2	1(26)	39	41	40	$-10(l_1 = 5)$	10	0.13	no
100	1.0d-20	off	69	71	70	$-100(l_1 = 5)$	$100(u_n = 10)$	0.08	no
100	1.0d-3	2(56)	58	114	59	$-100(l_1 = 5)$	$100(u_n = 10)$	0.11	no
1000	1.0d-20	off	43	93	44	$-10(l_1 = 5)$	10	0.23	no
1000	1.0d-4	1(36)	40	42	41	$-10(l_1 = 5)$	10	0.18	no
10000	1.0d-20	off	83	85	84	$-100(l_1 = 5)$	$100(u_n = 10)$	1.7	no
10000	1.0d-5	1(55)	77	127	78	$-100(l_1 = 5)$	$100(u_n = 10)$	2.37	no

Table 4: PSPG with and without preconditioning for Extended Powell Singular ($-0.5 < x_i^* < 0.5$ for all i) for different values of n and different values of l and u.

n	*tolpre*	Precond(t)	*iter*	F	G	l	u	Time	$x^* \in \Omega$
100	1.0d-20	off	336	566	337	$-1(l_1 = -10)$	$1000(u_1 = 30)$	0.14	yes
100	1.0d-1	1(31)	46	49	47	$-1(l_1 = -10)$	$1000(u_1 = 30)$	0.02	yes
1000	1.0d-20	off	322	581	323	$-1(l_1 = -10)$	$1000(u_1 = 30)$	0.95	yes
1000	1.0d-1	1(31)	46	49	47	$-1(l_1 = -10)$	$1000(u_1 = 30)$	0.12	yes
10000	1.0d-20	off	206	356	207	$-1(l_1 = -10)$	$1000(u_1 = 30)$	6.83	yes
10000	1.0d-1	1(31)	46	49	47	$-1(l_1 = -10)$	$1000(u_1 = 30)$	1.46	yes
100	1.0d-20	off	274	337	275	$-\infty$	0	0.09	no
100	1.0d-3	1(143)	157	222	158	$-\infty$	0	0.07	no
1000	1.0d-20	off	269	335	270	$-\infty$	0	0.66	no
1000	1.0d-3	1(142)	157	223	158	$-\infty$	0	0.41	no
10000	1.0d-20	off	269	335	270	$-\infty$	0	7.33	no
10000	1.0d-3	1(142)	157	222	158	$-\infty$	0	4.52	no

References

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

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

[3] E.G. Birgin, J.M. Martinez, and M. Raydan, Algorithm 813: SPG - software for convex-constrained optimization, *ACM Transactions on Mathematical Software*, **27** (2001), 340–349.

[4] E.G. Birgin, I. Chambouleyron, and J.M. Martinez, Estimation of the optical constants and the thickness of thin films using unconstrained optimization, *Journal of Computational Physics*, **151** (1999), 862–880.

[5] E.G. Birgin and Y.G. Evtushenko, Automatic differentiation and spectral projected gradient methods for optimal control problems, *Optimization Methods and Software*, **10** (1998), 125–146.

[6] E.G. Birgin, J.M. Martinez, and M. Raydan, Nonmonotone spectral projected gradient methods on convex sets, *SIAM J. Opt.*, **10** (2000), 1196–1211.

[7] Z. Castillo, D. Cores, and M. Raydan, Low cost optimization techniques for solving the nonlinear seismic reflection tomography problem, *Optimization and Engineering*, **1** (2000), 155–169.

[8] D. Cores, G. Fung, and R. Michelena, A fast and global two point low storage optimization technique for tracing rays in 2D and 3D isotropic media, *Journal of Applied Geophysics*, **45** (2000), 273–287.

[9] Y.H. Dai and R. Fletcher, On the asymptotic behaviour of some new gradient methods. Technical Report NA/212, Department of Mathematics, University of Dundee, Dundee, Scotland, 2003.

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

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

[12] W. Glunt, T.L. Hayden, and M. Raydan, Molecular conformations from distance matrices, *J. Comp. Chem.*, **14** (1993), 114–120.

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

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

[15] F. Luengo, M. Raydan, W. Glunt, and T.L. Hayden, Preconditioned spectral gradient method, *Numerical Algorithms*, **30** (2002), 241–258.

[16] B. Molina and M. Raydan, The preconditioned Barzilai-Borwein method for the numerical solution of partial differential equations. *Numerical Algorithms*, **13** (1996), 45–60.

[17] M. Raydan, On the Barzilai and Borwein choice of steplength for the gradient method, *IMA J. Numer. Anal.*, **13** (1993), 321–326.

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

[19] C.R. Vogel, A constrained least-squares regularization method for nonlinear ill-posed problems, *SIAM J. on Control and Optimization*, **28** (1990), 34–49.

[20] C. Wells, W. Glunt, and T.L. Hayden, Searching conformational space with the spectral distance geometry algorithm, *Journal of Molecular Structure (Theochem)*, **308** (1994), 263–271.