MAT

7

Primeras Jornadas

sobre Ecuaciones

Diferenciales,

Optimización y

Análisis Numérico

Primera Parte

MAT

SERIE A: CONFERENCIAS, SEMINARIOS Y TRABAJOS DE MATEMÁTICA

No. 7

PRIMERAS JORNADAS SOBRE ECUACIONES DIFERENCIALES, OPTIMIZACIÓN Y ANÁLISIS NUMÉRICO

Primera Parte

Domingo A. Tarzia (Ed.)

INDICE

Rosario, Octubre 2004

A Survey of the Spectral Gradient Method[1]

A.B. VERDIELL, M.C. MACIEL, L.S. OROFINO[2] and T.I. GIBELLI[3]

Departamento de Matemática, Universidad Nacional del Sur
Av. Alem 1253, 8000 Bahía Blanca, Argentina.

Abstract

Finite dimensional unconstrained minimization problems are generally solved by iterative methods like Newton, quasi-Newton, steepest descent, conjugate gradient methods and its variations. All of them share the same property: they are descent methods.

In 1988 Barzilai and Borwein introduce the spectral gradient method which have been analyzed by Raydán in 1991. Initially it has been developed for the quadratic case and later has been extended to the general unconstrained minimization problem by using a globalization strategy. Its main characteristic is the non-descent property. Furthermore the method requires few storage locations and very inexpensive computations.

The purpose of this work is to trace the development of the spectral gradient method, in particular when the problem is defined in a Hilbert space. The algorithms are treated in details but in dealing with them we have presented only the most general results available and we have given these a broad brush treatment.

Key words: Projected gradients, non-monotone line search, large scale problems, spectral gradient method.
AMS Subject Classification: 49M07, 49M10, 65K, 90C06, 90C20.

1 Introduction

Unconstrained minimization problems in finite dimension are generally solved by iterative methods like Newton, quasi-Newton, steepest descent and conjugate gradient methods. There are several variations of these methods and all of them share the same property: they are descent methods.

In 1991, Raydán [39, 40] introduces the spectral gradient method which is an extension to the Barzilai and Borwain method [3]. This method has been developed for the quadratic case and its main characteristic is the non-descent property. The method has also been extended to the general unconstrained minimization problem and it results to be globally convergent when a non-monotone line search is incorporated [41].

For the unconstrained problem defined in function space, methods like Newton and like conjugate gradients have been developed and well studied [2, 15, 16, 22, 23, 24, 26, 27, 28, 29, 30, 38, 44].

At the beginning of this decade some attempts to analyze the non-descent spectral gradient method for the infinite dimensional case have been done [1, 33].

The purpose of this work is to trace the development of the spectral gradient method, in particular when the problem is defined in a Hilbert space. The algorithms are treated in details but in dealing with them we have presented only the most general results available and we have

---

[2]{averdiel-immaciel-sorofino}@criba.edu.ar
[3]tgibelli@uns.edu.ar

given these a broad brush treatment.

This work is organized into sections, each dealing with a specific problem and the based-on-spectral gradient algorithm developed for them. Section 2 is devoted to the finite case. The third section to the infinite case and in the fourth the unconstrained control problem is described. The final remarks are established in section 5.

2   Finite dimensional case

Let be $f : \mathbb{R}^n \to \mathbb{R}$ a differentiable function. It is well known that any solution of the unconstrained minimization problem,

$$\min_{x \in \mathbb{R}^n} f(x) \tag{1}$$

solves the nonlinear equations problem:

$$\text{find } x_\star \in \mathbb{R}^n \text{ such that } \nabla f(x_\star) = 0. \tag{2}$$

The methods proposed to solve it are usually iterative procedures: if x_k denotes the current iterate, and if it is not a good estimator of $x_\star$, a better one, $x_{k+1} = x_k + s_k$ is required. Here s_k means the step and it can be obtained by different methods.

In many algorithms, each iteration involves the calculation of a quasi-Newton step: $s_k^{QN} = -A_k^{-1}\nabla f(x_k)$, where $A_k \in \mathbb{R}^{n \times n}$ is an approximation of the Hessian matrix of f at x_k. After each iteration, the current A_k is updated to A_{k+1}, and usually it is chosen satisfying the secant equation:

$$A_{k+1}s_k = y_k \tag{3}$$

where $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$.

If $n > 1$, this equation does not completely specify the matrix A_{k+1}, so in addition to obeying (3) there are some desirable properties for A_{k+1}. In the spectral gradient method the update is restricted to be $A_{k+1} = \alpha_{k+1}I$ where $\alpha_{k+1} \in \mathbb{R}$ solves the linear system $y_k = \alpha_{k+1}Is_k$ in the least square sense: that is, if $s_k \neq 0$

$$\alpha_{k+1} = \frac{s_k^T y_k}{s_k^T s_k}. \tag{4}$$

Taking these ideas into account, the Spectral Gradient algorithm (SG) is established as:

Algorithm 2.1 Given $x_0 \in \mathbb{R}^n, \alpha_0 \in \mathbb{R}$.
For $k = 0, 1, ...$, repeat until convergence:

Step 1. $s_k = -\dfrac{1}{\alpha_k}\nabla f(x_k)$

Step 2. $x_{k+1} = x_k + s_k$

Step 3. $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$

Step 4. $\alpha_{k+1} = \dfrac{s_k^T y_k}{s_k^T s_k}.$

It is important to point out that in this method, the new iterate is chosen in the same direction as the Cauchy method, but another step length is used and in numerical experiments the SG algorithm is significantly faster than the classic gradient method at the same cost per iteration.

The most important features of Algorithm 2.1 are:

- Every iteration requires two inner products, one scalar-vector multiplication, two vector additions and only one gradient evaluation.

- It is a gradient method which uses information of the two previous iterates. It makes a difference with the Cauchy method, which uses only information of the previous iterate.

- It satisfies the *weak* secant equation: $s_k^T A_{k+1} s_k = s_k^T y_k$.

- The scalar α_{k+1} is a Rayleigh quotient of the matrix

$$\int_0^1 \nabla^2 f(x_k + ts_k) dt.$$

If the objective function is quadratic,

$$f(x) = \frac{1}{2} x^T A x - b^T x + c \tag{5}$$

where A is a symmetric positive definite (SPD) matrix, (4) becomes in:

$$\alpha_{k+1} = \frac{s_k^T A s_k}{s_k^T s_k}. \tag{6}$$

In this case, **Steps 3-4** of the algorithm can be changed by (6) and because of α_{k+1} is the Rayleigh quotient of A at s_k, there is not danger of dividing by zero in **Step 1.**

There is a relationship between this method and the shifted power method to approximate eigenvalues and eigenvectors which is extensively used to establish local and global convergence results [39]. This relation holds for the infinite dimensional case and more details are shown in Section 2.

Barzilai and Borwein [3] have established r-superlinear rate of convergence if $n = 2$ and A has two distinct eigenvalues. The convergence of the method when applied to the minimization of a strictly convex quadratic function was established by Raydán [39, 40] in the following theorem:

Theorem 2.1 *Let f(x) be a strictly convex quadratic function. Let $\{x_k\}$ be the sequence generated by the Algorithm 2.1 and $x_\star$ the unique minimizer of f. Then, either $x_j = x_\star$ for some index j, or the sequence $\{x_k\}$ converges to $x_\star$.*

Related to the convergence rate, Raydán [39] proved the following results:

Theorem 2.2 *Let $f(x) = \frac{1}{2} x^T A x - b^T x + c$ where A is a SPD matrix that has only two distinct eigenvalues, $\lambda_1 < \lambda_2$. Assume that for same k, α_k is sufficiently close to λ_1 or λ_2. Then the sequence $\{x_k\}$ from Algorithm 2.1 converges to $x_\star$ q-quadratically. Furthermore, if α_k is equal to either λ_1 or λ_2 then, $x_{k+3} = x_\star$.*

Corollary 2.1 *Under the assumptions of Theorem 2.2, Algorithm 2.1 has the exact r-rate of convergence of $\sqrt[5]{4}$.*

In numerical experiments as well as in the convergence analysis, it can be seen one of the most important features of this proposal: it is a non-monotone method, in some cases the error's norm and the objective function value increase.

If $n > 2$, numerical experiments show that the performance of the algorithm depends on the eigenvalues of A and its condition number $\kappa(A)$. So, in order to improve convergence, this matrix is preconditioned in such a way that either $\kappa(A)$ is reduced or the eigenvalues of A are clustered.

In the quadratic case, the basic idea is to transform the objective function (5) in

$$\hat{f}(x) = \frac{1}{2}x^T \hat{A}x - \hat{b}^T x + \hat{c}$$

where $\hat{A} = E^{-1}AE^T$, $\hat{b} = E^{-1}b$, $\hat{c} = c$ for some nonsingular matrix E.

Therefore, if Algorithm 2.1 is applied to $\hat{f}(x)$ setting $C = EE^T$, $h_k = C^{-1}\nabla f(x_k)$, the preconditioned version is stated as follows:

Algorithm 2.2 Given $x_0 \in \mathbb{R}^n, \alpha_0 \in \mathbb{R}$ and C a symmetric and positive definite matrix, set $g_0 = Ax_0 - b$.
For $k = 0, 1, ...,$ repeat until convergence:

Step 1. Solve $Ch_k = g_k$ for h_k

Step 2. Set $p_k = Ah_k$

Step 3. Set $x_{k+1} = x_k - \dfrac{1}{\alpha_k}h_k$

Step 4. Set $g_{k+1} = g_k - \dfrac{1}{\alpha_k}p_k$

Step 5. Set $\alpha_{k+1} = \dfrac{h_k^T p_k}{g_k^T h_k}$.

The matrix C is called the *preconditioning matrix* and A the *preconditioned matrix*. Numerical results and comparison with the preconditioned conjugate gradient method can be found in [39].

For the general case, the method needs to be incorporated in a globalization scheme; Raydán proposes an algorithm based on the non-monotone line search strategy, introduced by Grippo, Lampariello y Lucidi [25] and proves global convergence [41] when the iterates are generated by the following algorithm.

Algorithm 2.3 Given $x_0 \in \mathbb{R}^n, \alpha_0 \in \mathbb{R}, \delta > 0, 0 < \sigma_1 < \sigma_2 < 1, \gamma \in (0,1), 0 < \epsilon < 1$ and $M \geq 0$, an integer. Set $k = 0, g_0 = \nabla f(x_0)$.

Step 1. If $\| g_k \| = 0$, stop.

Step 2. If $\alpha_k \leq \epsilon$ or $\alpha_k \geq \dfrac{1}{\epsilon}$, set $\alpha_k = \delta$

Step 3. Set $\lambda = \dfrac{1}{\alpha_k}$

Step 4. (non-monotone line search)
If $f(x_k - \lambda g_k) \leq \max\limits_{0 \leq j \leq min(k,M)} f(x_{k-j}) - \gamma\lambda g_k^T g_k$ set $\lambda_k = \lambda$, $x_{k+1} = x_k - \lambda_k g_k$ and go to Step 6.

Step 5. Choose $\sigma \in [\sigma_1, \sigma_2]$, set $\lambda = \sigma\lambda$ and go to Step 4.

Step 6. Set $\alpha_{k+1} = -\dfrac{g_k^T y_k}{\lambda_k g_k^T g_k}$, $k = k+1$ and go to Step 1.

Dai and Liao [14] have also established r-linear convergence of the method for any dimensional strictly convex quadratics and as a consequence of this result the method is also locally r-linear convergent for general objective functions, and therefore the stepsize in the SG method will always be accepted by the non-monotone line search when the iterate is close to the solution.

These results were extended for the minimization on convex sets [10]. Also, for the large scale non-linear optimization problem solved by a trust region strategy, at each iteration the quadratic subproblem is solved via an algorithm based on the algorithm designed for convex sets [32, 34].

3   The infinite case

In this section we analyze the behavior of the spectral gradient method when it is extended to the unconstrained quadratic problem defined in an infinite dimensional real Hilbert space H. Let A be a bounded operator defined on H and $\langle, \rangle$ the inner product on H. Given the quadratic functional $q(x) = \dfrac{1}{2}\langle x, Ax \rangle - \langle x, b \rangle$, let us consider the unconstrained minimization problem

$$\min_{x \in H} q(x). \tag{7}$$

Our objective is to analyze the convergence of the algorithm for a self-adjoint and strictly positive operator, which has a numerable amount of eigenvalues $\{\lambda_i\}$, its spectrum $\sigma(A)$ satisfies $\sigma(A) \subseteq [m, M]$ and the system of associate eigenvectors forms an orthonormal basis. Under these assumptions the Algorithm 1 can be sated on H as follows:

Algorithm 3.1 Given $x_0 \in H, \alpha_0 \in \mathbb{R}$,
For $k = 0, 1, ...,$ repeat until convergence:

Step 1. $s_k = -\dfrac{1}{\alpha_k}\nabla q(x_k)$

Step 2. $x_{k+1} = x_k + s_k$

Step 3. $y_k = \nabla q(x_{k+1}) - \nabla q(x_k)$

Step 4. $\alpha_{k+1} = \dfrac{\langle s_k, y_k \rangle}{\langle s_k, s_k \rangle}$.

Let us denote $g_k = \nabla q(x_k)$ the first Gâteaux derivative.
Using the relations among y_k, s_k and g_k results

$$\alpha_{k+1} = \frac{\langle g_k, Ag_k \rangle}{\langle g_k, g_k \rangle},$$

the Rayleigh quotient of A evaluated at g_k. Then

$$0 < m \leq \alpha_{k+1} \leq M. \tag{8}$$

Since the search direction is the negative gradient, the method belongs to the class of gradient methods. The next lemma shows the relationship with the power method what is essential to prove local and global convergence.

Lemma 3.1 *Let* $q(x) = \dfrac{1}{2}\langle x, Ax \rangle - \langle x, b \rangle$*, where* A *is self-adjoint and strictly positive. Let* x_* *be the unique minimizer of* q*,* $\{x_k\}$ *the sequence generated by the algorithm and* $e_k = x_\star - x_k$*, for all* k*. Then:*

1) $Ae_k = \alpha_k s_k = -g_k$

2) $e_{k+1} = \dfrac{1}{\alpha_k}(\alpha_k I - A)e_k$

3) $s_{k+1} = \dfrac{1}{\alpha_{k+1}}(\alpha_k I - A)s_k$

4) $g_{k+1} = \dfrac{1}{\alpha_k}(\alpha_k I - A)g_k.$

The relations established in Lemma 3.1 allow us to conclude that $||e_k||$ goes to zero if and only if $||s_k||$ goes to zero, and it is equivalent to prove that $||g_k||$ goes to zero.

For any initial error e_0 there are constants d_i^0 such that

$$e_0 = \sum_{i=1}^{\infty} d_i^0 v_i.$$

Again, the lemma 3.1 allows us to obtain the following expression for the error

$$e_{k+1} = \sum_{i=1}^{\infty} d_i^{k+1} v_i,$$

where

$$d_i^{k+1} = \left(\frac{\alpha_k - \lambda_i}{\alpha_k} \right) d_i^k = \prod_{j=0}^{k} \left(\frac{\alpha_j - \lambda_i}{\alpha_j} \right) d_i^0.$$

The convergence properties of the sequence $\{e_k\}$ will depend on the behavior of each of the sequences $\{d_i^k\}$. If the spectrum of A satisfies the condition $M < 2m$ we can establish the following result:

Lemma 3.2 *Let* $q(x) = \dfrac{1}{2}\langle x, Ax \rangle - \langle x, b \rangle$*, where* A *is strictly positive such that* $M < 2m$*. Let* $x_\star$ *the unique minimizer of* q*. Then, the sequence* $\{x_k\}$ *generated by the spectral gradient method converges* q*-linearly to* $x_\star$ *and the factor of convergence* $c < 1$ *is* $c = \dfrac{M-m}{m}.$

If the restrictive condition $M < 2m$ does not hold, the sequence $\{d_i^k\}$ show a non-monotone behavior, according to the relative position of the scalars α_k and the eigenvalues λ_i in the spectrum, however this situation does not perturb the convergence of the algorithm. Assuming that the sequence of eigenvalues λ_i is increasing

$$0 < m = \lambda_1 < \lambda_2 < \lambda_3......... \tag{9}$$

and $\lambda_i \to M$, the next lemmas help us to prove the main theorem.

Lemma 3.3 *The sequence* $\{d_1^k\}$ *converges to zero* q*-linearly and the factor of convergence is* $\hat{c}=1 - (m/M).$

Lemma 3.4 *If for a fixed integer l, $1 \leq l$, the sequences $\{d_1^k\},...,\{d_l^k\}$ converge to zero then*

$$\liminf_{k \to \infty} |d_{l+1}^k| = 0.$$

Theorem 3.1 *Let $q(x) = \dfrac{1}{2}\langle x, Ax \rangle - \langle x, b \rangle$, where A is self-adjoint and strictly positive operator and the sequence of eigenvalues satisfies (9). Let $\{x_k\}$ be the sequence generated by the spectral gradient method and $x_\star$ the unique minimizer of q. Then, either $x_j = x_\star$, for some j, or the whole sequence $\{x_k\}$ converges to $x_\star$.*

The proofs of these results can be found in [33].

4   An application to control problems

The spectral gradient method has successfully been used in different areas such as Geophysics [4, 6, 12, 13], Physics [7, 37], Chemistry [18, 19, 21, 20, 45], etc. Also have been developed algorithms to solve algebraic nonlinear systems [11, 31], partial differential equations [35, 42] and other nonlinear programming problems [5, 8, 9, 17, 36].

In this section we present an application of the spectral gradient method to an infinite dimension control problem. Optimal control problems belong to a more ample class: the differential equations-constrained optimization problem. A nice review of some aspects of PDE-constrained optimization can be found in [43]. Control problems and their discretized form, viewed as minimization problems have already been solved by quasi-Newton methods. Kelley and Sachs [29] analyze the behavior of the BFGS-secant method when it is applied to the control problem.

Let us consider the general nonlinear control problem to minimize

$$F(u) = \int_0^T L(x(t), u(t), t) dt \tag{10}$$

subject to

$$\begin{aligned} \dot{x} &= f(x(t), u(t), t) \\ x(0) &= x_0, \end{aligned} \tag{11}$$

where $L : \mathbb{R}^{n+m+1} \to \mathbb{R}$ and $f : \mathbb{R}^{n+m+1} \to \mathbb{R}$.
Recall, the gradient of F is given by

$$\nabla F(u) = p(.)^T f_u(x(.), u(.), .) + L_u(x(.), u(.), .), \tag{12}$$

and $p(t)$ solves the adjoint equation

$$\begin{aligned} -\dot{p} &= p(t)^T f_x(x(t)u(t), t) + L_x(x(t), u(t), t) \\ p(T) &= 0. \end{aligned} \tag{13}$$

The second derivative of $F(u)$ is given by

$$\begin{aligned} \langle w, \nabla^2 F(u)v \rangle &= \langle \xi(w), H_{xx}(x, u, .)\xi(v) \rangle + \langle w, H_{ux}(x, u, .)\xi(v) \rangle \\ &\quad + \langle \xi(w), H_{xu}(x, u, .)v \rangle, \end{aligned}$$

where $H(x, u, t) = p(t)f(x, u, t) + L(x, u, t)$, $\xi(w)$ solves an initial value problem and $\langle ., . \rangle$ denotes the inner product defined by $\langle u, v \rangle = \int_0^T u(t)v(t)dt$ for all $u, v \in L^2[0, T]$. It is clear that each

gradient evaluation involves the solution of a system of differential equations and it is necessary to compute $\nabla^2 F(u)$ if we want to apply Newton's method. These two facts suggest us to choose appropriate methods not only to solve the ordinary differential systems but also to approximate the Hessian matrix.

Let us consider the unconstrained problem

$$\min F(u), \qquad u \in H. \tag{14}$$

Recall that any solution of (14) is also solution of the nonlinear algebraic system $G(u) = \nabla F(u) = 0$.

The Hilbert space H is approximated by a finite dimensional space H^N and let us replace the functional F by the functional F^N defined on H^N and finally we consider the minimization problem

$$\min F^N(u^N), \qquad u^N \in H^N. \tag{15}$$

Any solution of (15) is also solution of $G^N(u^N) = \nabla F^N(u^N) = 0$.

The algorithm will be analyzed under the following assumptions.

Let $\{P^N\}$ denote a sequence of linear prolongation operators

$$P^N : H^N \to Z,$$

being Z a normed subspace of H, with the property that $\|.\|_Z \geq \|.\|_H$, where $\|.\|_H$ and $\|.\|_Z$ denote the norms in the Hilbert space H and the subspace Z respectively.

Let $\langle , \rangle_N$ denote the inner product on H^N and $\|.\|_N$ the induced norm by such inner product. Let $G^N : H^N \to H^N$ be a Fréchet-differentiable operator.

A sequence $u^N \in H^N$ is said to be Z-convergent to u, and we denote $u^N \underset{Z}{\to} u \in Z$, if

$$\lim_{N \to \infty} \|P^N u^N - u\|_Z = 0.$$

Assuming,

A1) If $u^N \underset{Z}{\to} u \in Z$, $v^N \underset{Z}{\to} v \in Z$ then $\lim_{N \to \infty} \langle u^N, v^N \rangle_N = \langle u, v \rangle$.

A2) If $u^N \underset{Z}{\to} u$ then $G^N(u^N) \underset{Z}{\to} G(u)$

A3) If $u^N \underset{Z}{\to} u$ then $\lim_{N \to \infty} F^N(u^N) = F(u)$.

The assumptions A1) and A2) are the same as the established by Kelley and Sachs [29]. In order to analyze the spectral gradient algorithm we need the assumption A3).

The discretized function $F^N : H^N \to \mathbb{R}$ from the control problem is not strictly convex. However, it is interesting to analyze the behavior of the algorithm when F^N is strictly convex because of the properties of the finite dimensional algorithm.

We state the following results, whose proof can be found in [1].

Theorem 4.1 *Let $G : H \to H$ and $G^N : H^N \to H^N$ be under the assumptions (A1) and (A2). If for all N, the sequence $\{u_k^N\}_{k \in \mathbb{N}}$ satisfies:*

i) For all $N \in \mathbb{N}$, $\lim_{k \to \infty} \|G^N(u_k^N)\|_N = 0$.

ii) For all $k \in \mathbb{N}$, $u_k^N \underset{Z}{\rightarrow} u_k \in Z$,

then $\lim\limits_{k \to \infty} \|G(u_k)\|_H = 0$.

Lemma 4.1 *Let $G : H \to H$ and $G^N : H^N \to H^N$ be under assumptions (A1) and (A2). If $u_0 \in H$ and $\alpha_0 \in \mathbb{R}$, let the sequence $\{u_0^N\}_{N \in \mathbb{N}}$ satisfies $u_0^N \underset{Z}{\rightarrow} u_0$ and $\alpha_0^N = \alpha_0$ for all N. If $\{u_k^N\}_{k \in \mathbb{N}}$ is the sequence generated by algorithm 2.1, starting with u_0^N and α_0^N for each $N \in \mathbb{N}$, then for all $k \in \mathbb{N}$,*

$$u_k^N \underset{Z}{\rightarrow} u_k, \quad \alpha_k^N \underset{N \to \infty}{\longrightarrow} \alpha_k$$

Corollary 4.1 *Let $G : H \to H$ and $G^N : H^N \to H^N$ be under the conditions (A1) and (A2) and assume that F^N satisfies the conditions required by the convergence theorem of the spectral gradient method applied to the finite case. Let $N \in \mathbb{N}$, $\{u_k^N\}_{k \in \mathbb{N}} \subset H^N$ be the sequence generated by algorithm 2.1. Then*

a) For all $N \in \mathbb{N}$, $\lim\limits_{k \to \infty} \|G^N(u_k^N)\|_N = 0$.

b) For all $k \in \mathbb{N}$, $u_k^N \underset{Z}{\rightarrow} u_k \in Z$.

From the corollary it is possible to conclude that the sequence $\{u_k^N\}$ generated by the Algorithm 2.1 satisfies the assumptions of Theorem 4.1. It means that the sequence generated by the algorithm for the finite dimensional case approximates to the solution of the infinite dimensional problem for a discretization with N large enough.

If the functional $F^N : H^N \to \mathbb{R}$ is a non-linear function not necessarily strictly convex, a globalization strategy is incorporated to the algorithm [41]. Even though global convergence results are known for this case, it is not possible to extend the proof to the infinite dimensional case. When the BFGS method is applied, Kelley and Sachs [29] use the rate of convergence of the algorithm in Hilbert space.

If the globalized spectral gradient method is applied, the difficulty appears when the convergence of $\{\alpha_k^N\}$ to α_k has to be proved. In this case the value α_k^N varies when the line search is used. The number of inner iterations requires by the line search can change with the dimension of the problem. This argument is illustrated in the tables.

Since in the optimal control problem the Hilbert space is $H = L^2[0, T]$, let $N \in \mathbb{N}$ be the discretization parameter, the time interval $[0, T]$ is divided into N subinterval of equal length $h = \frac{T}{N}$. The approximate finite space is $H^N = \mathbb{R}^N$, with the inner product $\langle, \rangle_N$ defined by the composite Simpson's rule.

At each iteration the state and the adjoint equations must be solved to evaluate the functions F^N and G^N. The solutions $x(t)$ and $p(t)$ of the differential equations (11) y (13) are approximated by the fourth order Runge Kutta and Hermite interpolation. The evaluation of $F^N(u^N)$ is made by the composite Simpson's rule. The function $G^N = p^{N^T} f_x^N + L_x^N$ approximates the gradient G in the equation (12) is not, in general, the gradient of the scalar valued function F^N. The following lemma shows that, in spite of this, the directions chosen are descent directions.

Lemma 4.2 *Assuming the assumptions (A2) and (A3), let $G^N = p^{N^T} f_x^N + L_x^N$ be a discretization of G, the gradient of $F(u)$. Let $\{u^N\}_{N \in \mathbb{N}}$ be a sequence such that $u^N \underset{Z}{\rightarrow} u$. If $\|G^N(u^N)\| \neq 0$ then $-G^N(u^N)$ is a descent direction of F^N from u^N, for all $N \geq N_0$, for some $N_0 \geq 0$.*

With these choices of the inner product, the functions F^N and G^N, the assumptions (A1)-(A3) are verified.

Different problems have been tested with different discretizations [1]. The low cost of the spectral gradient method allowed us to use high discretizations obtaining very good approximations to the solutions. The following example represents a simple production inventory model. It has beeen solved by Kelley and Sachs [29] by using the BFGS secant method.
The description is the following:

$$F(u) = \int_0^T \frac{e^{-\rho t}}{2} \left(d(x - a(t))^2 + c(u - b(t))^2\right) dt \tag{16}$$

subject to

$$\dot{x} = u(t) - s(t)$$
$$x(0) = x_0.$$

The parameters were chosen as follows:

$$T = 0.3, \rho = 1, d = c = 1, a = 15, b = 30, s(t) = t^2, x_0 = 10.$$

Using Pontryagin maximum principle, it is possible to compute the optimal control $u_\star$ for the infinite dimensional problem. It is of the form

$$u_\star(t) = \alpha_1 \lambda_1 e^{\lambda_1 t} + \alpha_2 \lambda_2 e^{\lambda_2 t} + t^2 - 2t + 4$$

where $\lambda_{1/2} = (1 \pm \sqrt{5})/2$ and α_i are other constants. More details about them can be found in [29].

Following the framework proposed by Kelly and Sachs, we use different discretization to solve the problems. The following table report the number of iterates necessary to achieve the tolerance in the norm of the gradient, for different values of N.

Number of iterations		
N	$u_0 = u_\star + 100$	$u_0 = \sqrt{u_\star} + 100$
400	14	13
800	14	13
1600	15	13
3200	15	14
6400	16	16
12800	18	18
25600	26	23

We have already said that the inductive proof of the convergence of the algorithm does not hold when the non monotone line search is added. It occurs because the number of line searches per iteration is independent of the discretization. This fact can be observed in the following table where each iteration of problem is detailed by using a discretization with $N = 12800$ and starting point $u_0 = u_\star + 100$.

Iter	F	$\|G^N\|$	LS
1	1309.91610867154	102.660152979672	3
2	1005.86083986541	87.9911473130837	3
3	772.951943556889	75.4104396549632	2
4	438.939972885863	53.7843562176199	1
5	114.728359380689	27.6994895434872	0
6	2.05149541312622	4.75685850055365	0
7	0.951154993329865	0.745660189477803	0
8	0.934393423093945	0.110252930849182	0
9	0.934121964482021	2.149098455769760E-002	0
10	0.934115833756604	2.579978101994840E-003	0
11	0.934115741378638	4.558961104628168E-004	0
12	0.934115739509478	8.442761702198709E-005	0
13	0.934115739453236	4.945139366441254E-006	0
14	0.934115739453054	1.620860793138945E-006	0
15	0.934115739453047	1.300917098490117E-007	0
16	0.934115739453041	4.037641332388375E-009	0
17	0.934115739453054	1.043105157805257E-009	0
18	0.934115739453049	9.858780458671390E-012	

The following table shows the number of line search at each iteration para first iterations of problem, starting with $u_0 = exp(u_\star + 100)$, and using 11 different discretizations.

Iterate n^o	Number of points in the discretization										
	1600	3200	6400	12800	16000	19200	22400	24200	25000	25600	26000
1	99	100	102	103	103	104	104	104	104	104	104
2	0	0	1	3	3	4	4	4	4	5	5
3	0	0	0	3	3	4	4	4	4	4	4
4	0	0	0	2	2	4	4	4	4	4	4
5	0	0	0	1	1	3	4	4	4	4	4
6	0	0	0	0	0	3	3	3	4	4	4
7	0	0	0	0	0	2	3	3	3	4	4
8	0	0	0	0	0	1	2	3	3	3	3
9	0	0	0	0	0	0	2	2	2	3	3
10	0	0	0	0	0	0	0	1	2	2	2
11	0	0	0	0	0	0	0	0	1	2	2
12	0	0	0	0	0	0	0	0	0	1	1
13	0	0	0	0	0	0	0	0	0	0	0
$\vdots$	$\vdots$	$\vdots$									$\vdots$
	$\vdots$	$\vdots$									$\vdots$
Total iter.	64	64	67	68	68	72	72	72	74	75	75

5   Conclusions

We have presented a survey of the spectral gradient method for the unconstrained case. It is a relatively novel non-descent method, appropriate to large scale optimization problems and competitive with the traditional conjugate gradient method.

The extension of the method to an infinite Hilbert space has been analyzed for the quadratic case when the Hessian operator is self-adjoint. In this case the extension is straightforward. The compact operator case is currently being analyzed.

It is important to remark that for the infinite dimensional nonlinear case was not possible to

prove the convergence of the method because the global convergence results for the finite case can not be extended. In spite of this fact, the algorithm has been applied to a well known control problem and the numerical results are much better that the obtained by a secant method. These facts encourage us to continue analyzing the spectral gradient method for the infinite dimensional optimization problems.

References

[1] J.I. ARDENGHI, M.C. MACIEL, and T.I. GIBELLI. The spectral gradient method for unconstrained optimal control problems. Technical Report Submitted to IMA Journal of Numerical Analysis, Department of Mathematic, Southern National University, Av. Alem 1253, 8000 Bahía Blanca, Argentina, 2003.

[2] O. AXELSSON and J. KARATSON. On the rate of convergence of the conjugate gradient method for linear operators in Hilbert space. *Numerical Functional Analysis and Optimization*, 23 (3&4):285–302, 2002.

[3] J. BARZILAI and J.M. BORWEIN. Two point step size gradient methods. *IMA Journal of Numerical Analysis*, 8:141–148, 1988.

[4] L. BELLO and M. RAYDAN. Convex constrained optimization for the seismic reflection tomography problem, 2002. Submitted to Optimization and Enineering.

[5] L. BELLO and M. RAYDAN. Preconditioned spectral projected gradient method on convex sets, 2002. Submitted to Journal of Computational Mathematics

[6] E.G. BIRGIN, R. BILOTI, M. TYGEL, and L.T. SANTOS. Restricted optimization: a clue to fast and accurate implementation of the common reflection surface method. *Journal of Applied Geophysics*, 42:143–155, 1999.

[7] 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*, 151:862–880, 1999.

[8] E.G. BIRGIN and J.M. MARTINEZ. A spectral conjugate gradient method for unconstrained optimization. *Applied Mathematics and Optimization*, 43:117–128, 2001.

[9] E.G. BIRGIN and J.M. MARTINEZ. Large-scale active-set box-constrained optimization method with spectral projected gradients. *Computational Optimization and Applications*, 23:101–125, 2002.

[10] E.G. BIRGIN, J.M. MARTINEZ, and M. RAYDAN. Nonmonotone spectral projected gradient methods on convex sets. *SIAM Journal on Optimization*, 10(4):1196–1211, 2000.

[11] C. BREZINSKI and J. P. CHEHAB. Multiparameter iterative schemes for the solution of systems of linear and nonlinear equations. *SIAM Journal on Scientific Computing*, 20:2140–2159, 1999.

[12] Z. CASTILLO, D. CORES, and M. RAYDAN. Slow cost optimization techniques for solving the nonlinear seismic reflection tomography problem. *Optimization and Enineering*, 1:155–169, 2000.

[13] 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*, 45:273–287, 2000.

[14] Y-H. DAI and L-Z. LIAO. *r*-linear convergence of the Barzilai and Borwein gradient method. *IMA Journal on Numerical Analysis*, 22:1–10, 2002.

[15] J.W. DANIEL. The conjugate gradient method for linear and nonlinear operator equations. *SIAM Journal on Numerical Analysis*, 14:10–26, 1967.

[16] J.W. DANIEL. Applications and methods for the minimization of functionals. In L.B. RALL, editor, *Nonlinear functional analysis and applications*, pages 399–424. Academic Press, New York, 1971.

[17] M. A. DINIZ-EHRHARDT, M. A. GOMES-RUGGIERO, J. M. MARTINEZ, and S. A. SANTOS. Augmented Lagrangian algorithms based on the spectral gradient for solving nonlinear programming problems, 2000.

[18] W. GLUNT, T.L. HAYDEN, and W. LIU. The embedding problem for predistance matrices. *Bulletin of Mathematical Biology*, 53:769–796, 1991.

[19] W. GLUNT, T.L. HAYDEN, and M. RAYDAN. Molecular conformations from distance matrices. *Journal of Computational Chemistry*, 14:114–120, 1993.

[20] W. GLUNT, T.L. HAYDEN, and M. RAYDAN. Preconditioners for distance matrix algorithms. *Journal of Computational Chemistry*, 15:227–232, 1994.

[21] W. GLUNT, T.L. HAYDEN, C. WELLS, J.G. SHELLING, and D.J. WARD. Applications of weighting and chirality strategies for distance geometry algorithms to an enterotoxin peptide analog. *Journal of Mathematical Chemistry*, 15:353–366, 1994.

[22] A.A. GOLDSTEIN. On steepest descent. *Journal of SIAM, Serie A, Control*, 3(1):147–151, 1965.

[23] A. GRIEWANK. Rates of convergence for secant methods on nonlinear problems in Hilbert space. In J.P. Hennart, editor, *Lectures Notes in Mathematics, # 1230*, pages 138–157. Springer-Verlag, Heidelberg, 1986.

[24] A. GRIEWANK. The local convergence of Broyden-like methods on Lipschitzian problems in Hilbert spaces. *SIAM Journal on Numerical Analysis*, 24(3):684–705, 1987.

[25] L. GRIPPO, F. LAMPARIELLO, and S. LUCIDI. A nonmonotone line search technique for Newton's method. *SIAM Journal on Numerical Analysis*, 23:707–716, 1986.

[26] R.M. HAYES. Iterative methods for solving linear problems in Hilbert spaces. *National Bureau of Standards Appl. Math. Ser.*, 39:71–104, 1954.

[27] L.B. HORWITZ and P.E. SARACHIK. Davidon's method in Hilbert spaces. *SIAM Journal on Applied Mathematics*, 16(4):676–695, 1968.

[28] C.T. KELLEY and E.W. SACHS. A quasi-Newton method for elliptic boundary value problems. *SIAM Journal on Numerical Analysis*, 24(3):516–531, 1987.

[29] C.T. KELLEY and E.W. SACHS. Quasi-Newton methods and unconstrained optimal control problems. *SIAM Journal on Control and Optimization*, 25(6):1503–1516, 1987.

[30] C.T. KELLEY and E.W. SACHS. Approximate quasi-Newton methods. *Mathematical Programming*, 48:41–70, 1990.

[31] W. LA-CRUZ and M. RAYDAN. Nonmonotone spectral methods for large-scale nonlinear systems. *Optimization Methods and Software*, 18:583–599, 2003.

[32] M.C. MACIEL, J.M. MARTINEZ, M. RAYDAN, and A.B. VERDIELL. A trust-region algorithm based on the spectral projected gradient method. In preparation.

[33] M.C. MACIEL and S.L. OROFINO. The spectral gradient method on Hilbert spaces. Technical Report Submitted to Numerical Functional Analysis and Optimization, Department of Mathematic, Southern National University, Av. Alem 1253, 8000 Bahía Blanca, Argentina, 2003.

[34] M.C. MACIEL, E.A. PILOTTA, and A.B. VERDIELL. A trust-region algorithm based on the spectral projected gradient method for nonlinear programming. In preparation.

[35] B. MOLINA and M. RAYDAN. Preconditioned Barzilai-Borwein method for the numerical solution of partial differential equations. *Numerical Algorithms*, 13:45–60, 1996.

[36] B. MOLINA and M. RAYDAN. Spectral variants of Krylov subspace methods. *Numerical Algorithms*, 29:197–208, 2002.

[37] M. MULATO, I. CHAMBOULEYRON, E.G. BIRGIN, and J.M. MARTINEZ. Determination of thickness and optical constants of a-Si:H films from transmittance data. *Applied Physics Letters*, 77:2133–2135, 2000.

[38] E. POLAK. An historical survey of computational methods in optimal control. *SIAM review*, 15(2):553–584, 1973.

[39] M. RAYDAN. *Convergence properties of the Barzilai and Borwein gradient method*. PhD thesis, Dept. of Mathematical Science, Rice University, Houston, Texas, 1991.

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

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

[42] M. RAYDAN. Preconditioned spectral gradient for solving nonlinear Poisson-type equations. In M. Cerrolaza C. Muller-Karger, M. Lentini, editor, *Proceedings CIMENICS 2002*, pages 47–55. Editorial USB, Caracas, Venezuela, 2002.

[43] E.W. SACHS. PDE constrained optimization. *SIAG/OPT Views-and-News*, 14(1):7–10, 2003.

[44] J. STOER. Two examples on the convergence of certain rank-2 minimization methods. *Linear Algebra and its Applications*, 28:217–222, 1979.

[45] C. WELLS, W. GLUNT, and T.L. HAYDEN. Searching conformational space with the spectral distance geometry algorithm. *Journal of Molecular Structure (Theochem)*, 308:263–271, 1994.