Journal of Statistical Software

MMMMMM YYYY, Volume VV, Issue II. http://www.jstatsoft.org/

Spectral Projected Gradient methods: Review and Perspectives

E. G. Birgin
University of São Paulo

J. M. Martínez
University of Campinas

M. Raydan
Universidad Simón Bolívar

Abstract

Over the last two decades, it has been observed that using the gradient vector as a search direction in large-scale optimization may lead to efficient algorithms. The effectiveness relies on choosing the step lengths according to novel ideas that are related to the spectrum of the underlying local Hessian rather than related to the standard decrease in the objective function. A review of these so-called spectral projected gradient methods for convex constrained optimization is presented. To illustrate the performance of these low-cost schemes, an optimization problem on the set of positive definite matrices is described.

Keywords: Spectral Projected Gradient methods, nonmonotone line search, large scale problems, convex constrained problems.

1. Introduction

A pioneering paper by Barzilai and Borwein (1988) proposed a gradient method for the unconstrained minimization of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ that uses a novel and nonstandard strategy for choosing the step length. Starting from a given $x^0 \in \mathbb{R}^n$, the Barzilai-Borwein (BB) iteration is given by

$$x^{k+1} = x^k - \lambda_k \nabla f(x^k), \tag{1}$$

where the initial step length $\lambda_0 > 0$ is arbitrary and, for all $k = 1, 2, \ldots$,

$$\lambda_k = \frac{s_{k-1}^\top s_{k-1}}{s_{k-1}^\top y_{k-1}}, \tag{2}$$

where $s_{k-1} = x^k - x^{k-1}$ and $y_{k-1} = \nabla f(x^k) - \nabla f(x^{k-1})$.

When $f(x) = \frac{1}{2}x^\top Ax + b^\top x + c$ is a quadratic function and A is a symmetric positive definite (SPD) matrix, then the step length (2) becomes

$$\lambda_k = \frac{\nabla f(x^{k-1})^\top \nabla f(x^{k-1})}{\nabla f(x^{k-1})^\top A \nabla f(x^{k-1})}. \tag{3}$$

Curiously, the step length (3) used in the BB method for defining x^{k+1} is the one used in the optimal Cauchy steepest descent method (Cauchy 1847) for defining the step at iteration k. Therefore, the BB method computes, at each iteration, the step that minimizes the quadratic objective function along the negative gradient direction but, instead of using this step at the k-th iteration, saves the step to be used in the next iteration. The main result in the paper by Barzilai and Borwein (1988) is to show the surprising result that for two-dimensional strictly convex quadratic functions the BB method converges R-superlinearly, which means that, in the average, the quotient between the errors at consecutive iterations tends asymptotically to zero. This fact raised the question about the existence of a gradient method with superlinear convergence in the general case. In 1990, the possibility of obtaining superlinear convergence for arbitrary n was discarded by Fletcher (1990), who also conjectured that, in general, only R-linear convergence should be expected. For more information about convergence rates see (Dennis and Schnabel 1983). Raydan (1993) established global convergence of the BB method for the strictly convex quadratic case with any number of variables. The nonstandard analysis presented in (Raydan 1993) was later refined by Dai and Liao (2002) to prove the expected R-linear convergence result, and it was extended by Friedlander, Martínez, Molina, and Raydan (1999) to consider a wide variety of delayed choices of step length for the gradient method; see also (Raydan and Svaiter 2002).

For the minimization of general (not necessarily quadratic) functions, a bizarre behavior seemed to discourage the application of the BB method: the sequence of functional values $f(x^k)$ did not decrease monotonically and, sometimes, monotonicity was severely violated. Nevertheless, it was experimentally observed (see, e.g., (Fletcher 1990; Glunt, Hayden, and Raydan 1993)) that the potential effectiveness of the BB method was related to the relationship between the step lengths and the eigenvalues of the Hessian rather than to the decrease of the function value. It was also observed (see, e.g., (Molina and Raydan 1996)) that, with the exception of very special cases, the BB method was not competitive with the classical Hestenes-Stieffel (Hestenes and Stiefel 1952) Conjugate Gradients (CG) algorithm when applied to strictly convex quadratic functions. On the other hand, starting with the work by Grippo, Lampariello, and Lucidi (1986), nonmonotone strategies for function minimization began to become popular when combined with Newton-type schemes. These strategies made it possible to define globally convergent algorithms without monotone decrease requirements. The philosophy behind nonmonotone strategies is that, frequently, 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.

The conditions were given for the implementation of the BB method for general unconstrained minimization with the help of a nonmonotone strategy. Raydan (1997) developed a globally convergent method in 1997 using the Grippo-Lampariello-Lucidi (GLL) strategy (Grippo *et al.* 1986) and the BB method given by (1) and (2). He 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 method of choice for solving strictly convex quadratic problems, its efficiency is hardly inherited by generalizations for minimizing general functions. Therefore, there existed a wide space for variations and extensions of the BB original method.

The Spectral Projected Gradient (SPG) method (Birgin, Martínez, and Raydan 2000, 2001, 2003b), for solving convex constrained problems, was born from the marriage of the global Barzila-Borwein (spectral) nonmonotone scheme (Raydan 1997) with the classical projected gradient (PG) method (Bertsekas 1976; Goldstein 1964; Levitin and Polyak 1966) which have been extensively used in statistics; see e.g., (Bernaards and Jennrich 2005) and references in there. Indeed, the effectiveness of the classical PG method can be greatly improved by incorporating the spectral step length and nonmonotone globalization strategies; more details on this topic can be found in (Bertsekas 1999). In Section 2, we review the basic idea of the SPG method and list some of its most important properties. In Section 3, we illustrate the properties of the method by solving an optimization problem on the convex set of positive definite matrices. In Section 4, we briefly describe some of the most relevant applications and extensions of the SPG method. Finally, in Section 5, we present some conclusions.

2. Spectral Projected Gradient (SPG) method

Quasi-Newton secant methods for unconstrained optimization (Dennis and Moré 1977; Dennis and Schnabel 1983) obey the recursive formula

$$x^{k+1} = x^k - \alpha_k B_k^{-1} \nabla f(x^k), \tag{4}$$

where the sequence of matrices $\{B_k\}$ satisfies the *secant equation*

$$B_{k+1} s_k = y_k. \tag{5}$$

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 are looking for a matrix B_{k+1} with a very simple structure that satisfies (5). More precisely, if we impose that

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

with $\sigma_{k+1} \in \mathbb{R}$, then equation (5) becomes:

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

In general, this equation has no solutions. However, accepting the least-squares solution that minimizes $\|\sigma s_k - y_k\|_2^2$, we obtain:

$$\sigma_{k+1} = \frac{s_k^\top y_k}{s_k^\top s_k}, \tag{6}$$

i.e., $\sigma_{k+1} = 1/\lambda_{k+1}$, where λ_{k+1} is the BB choice of step length given by (2). Namely, the method for unconstrained minimization is of the form $x^{k+1} = x^k + \alpha_k d_k$, where at each iteration,

$$d_k = -\lambda_k \nabla f(x^k),$$

$\lambda_k = 1/\sigma_k$, and formula (6) 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 $0 < \lambda_{\min} \leq \lambda_{\max} < \infty$ and defines, at each iteration:

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

By the Mean-Value Theorem of integral calculus, one has:

$$y_k = \left[\int_0^1 \nabla^2 f(x^k + ts_k)\, dt \right] s_k.$$

Therefore, (6) defines a Rayleigh quotient relative to the average Hessian matrix $\int_0^1 \nabla^2 f(x_k + ts_k)\, dt$. This coefficient is between the minimum and the maximum eigenvalue of the average Hessian, which motivates the denomination of *spectral method* (Birgin *et al.* 2000).

Writing the secant equation as $H_{k+1} y_k = s_k$, which is also standard in the Quasi-Newton tradition, we arrive at a different spectral coefficient: $(y_k^\top y_k)/(s_k^\top y_k)$; see (Barzilai and Borwein 1988; Raydan 1993). Both this dual and the primal (6) spectral choices of step lengths produce fast and effective nonmonotone gradient methods for large-scale unconstrained optimization (Fletcher 2005; Friedlander *et al.* 1999; Raydan 1997). Fletcher (2005) presents some experimental considerations about the relationship between the nonmonotonicity of BB-type methods and their surprising computational performance; pointing out that the effectiveness of the approach is related to the eigenvalues of the Hessian rather than to the decrease of the function value; see also (Asmundis, Serafino, Riccio, and Toraldo 2012; Glunt *et al.* 1993; Raydan and Svaiter 2002). A deeper analysis of the asymptotic behavior of BB methods and related methods is presented in (Dai and Fletcher 2005). The behavior of BB methods has also been analyzed using chaotic dynamical systems (van den Doel and Ascher 2012). Moreover, in the quadratic case, several spectral step lengths can be interpreted by means of a simple geometric object: the Bézier parabola (Berlinet and Roland 2011). All of these interesting theoretical as well as experimental observations, concerning the behavior of BB methods for unconstrained optimization, justify the interest in designing effective spectral gradient methods for constrained optimization.

The SPG method (Birgin *et al.* 2000, 2001, 2003b) is the spectral option for solving convex constrained optimization problems. As its unconstrained counterpart (Raydan 1997), the SPG method has the form

$$x^{k+1} = x^k + \alpha_k d_k, \tag{7}$$

where the search direction d_k has been defined in (Birgin *et al.* 2000) as

$$d_k = P_\Omega(x^k - \lambda_k \nabla f(x^k)) - x^k,$$

P_Ω denotes the Euclidean projection onto the closed and convex set Ω, and λ_k is the spectral choice of step length (2). A related method with approximate projections has been defined in (Birgin *et al.* 2003b). The feasible direction d_k is a descent direction, i.e., $d_k^\top \nabla f(x^k) < 0$ which implies that $f(x^k + \alpha d_k) < f(x^k)$ for α small enough. This means that, in principle, one could define convergent methods imposing sufficient decrease at every iteration. However, as in the unconstrained case, this leads to very inefficient practical results. A key feature is to accept the initial BB-type step length as frequently as possible while simultaneously guarantee

global convergence. For this reason, the SPG method employs a nonmonotone line search that does not impose functional decrease at every iteration. In (Birgin *et al.* 2000, 2001, 2003b) the nonmonotone GLL (Grippo *et al.* 1986) search is used (see Algorithm 2.2 below). The global convergence of the SPG method, and some related extensions, can be found in (Birgin *et al.* 2003b).

The nonmonotone sufficient decrease criterion, used in the SPG method, 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 $\gamma \in (0, 1)$. Algorithm 2.1 describes the SPG method in details, while Algorithm 2.2 describes the line search. More details can be found in (Birgin *et al.* 2000, 2001).

Algorithm 2.1: Spectral Projected Gradient

Assume that a sufficient decrease parameter $\gamma \in (0, 1)$, an integer parameter $M \geq 1$ for the nonmonotone line search, safeguarding parameters $0 < \sigma_1 < \sigma_2 < 1$ for the quadratic interpolation, safeguarding parameters $0 < \lambda_{\min} \leq \lambda_{\max} < \infty$ for the spectral step length, an arbitrary initial point x^0, and $\lambda_0 \in [\lambda_{\min}, \lambda_{\max}]$ are given. If $x^0 \notin \Omega$ then redefine $x^0 = P_\Omega(x^0)$. Set $k \leftarrow 0$.

Step 1. *Stopping criterion*

If $\|P_\Omega(x^k - \nabla f(x^k)) - x^k\|_\infty \leq \varepsilon$ then stop declaring that x^k is an approximate stationary point.

Step 2. *Iterate*

Compute the search direction $d_k = P_\Omega(x^k - \lambda_k \nabla f(x^k)) - x^k$, compute the step length α_k using Algorithm 2.2 (with parameters γ, M, σ_1, and σ_2), and set $x^{k+1} = x^k + \alpha_k d_k$.

Step 3. *Compute the spectral step length*

Compute $s_k = x^{k+1} - x^k$ and $y_k = \nabla f(x^{k+1}) - \nabla f(x^k)$. If $s_k^\top y_k \leq 0$ then set $\lambda_{k+1} = \lambda_{\max}$. Otherwise, set $\lambda_{k+1} = \max\{\lambda_{\min}, \min\{s_k^\top s_k / s_k^\top y_k, \lambda_{\max}\}\}$. Set $k \leftarrow k + 1$ and go to Step 1.

Algorithm 2.2: Nonmonotone line search

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

Step 1. *Test nonmonotone GLL criterion*

If $f(x^k + \alpha d_k) \leq f_{\max} + \gamma \alpha \nabla f(x^k)^\top d_k$ then set $\alpha_k \leftarrow \alpha$ and stop.

Step 2. *Compute a safeguarded new trial step length*

Compute $\alpha_{\rm tmp} \leftarrow -\frac{1}{2}\alpha^2 \nabla f(x^k)^\top d_k / [f(x^k + \alpha d_k) - f(x^k) - \alpha \nabla f(x^k)^\top d_k]$. If $\alpha_{\rm tmp} \in [\sigma_1, \sigma_2 \alpha]$ then set $\alpha \leftarrow \alpha_{\rm tmp}$. Otherwise, set $\alpha \leftarrow \alpha/2$. Go to Step 1.

In practice, it is usual to set $\lambda_{\min} = 10^{-30}$, $\lambda_{\max} = 10^{30}$, $\sigma_1 = 0.1$, $\sigma_2 = 0.9$, and $\gamma = 10^{-4}$. A typical value for the nonmonotone parameter is $M = 10$, although in some applications values ranging from 2 to 100 have been reported. However, the best possible value of M is problem-dependent and a fine tuning may be adequate for specific applications. It can only be said that $M = 1$ is not good because it makes the strategy to coincide with the classical monotone Armijo strategy (a comparison with this case can be found in (Birgin *et al.* 2000)) and that $M \approx \infty$ is not good either because in this case the decrease of the objective function is not controlled at all. Other nonmonotone techniques were also introduced in (Grippo, Lampariello, and Lucidi 1991; Toint 1996; Dai and Zhang 2001; Zhang and Hager 2004; Grippo and Sciandrone 2002; La Cruz, Martínez, and Raydan 2006). Parameter $\lambda_0 \in [\lambda_{\min}, \lambda_{\max}]$ is arbitrary. In (Birgin *et al.* 2000, 2001) it was considered

$$\lambda_0 = \max\{\lambda_{\min}, \min\{1/\|P_\Omega(x^0 - \nabla f(x^0)) - x^0\|_\infty, \lambda_{\max}\}\}, \tag{8}$$

assuming that x^0 is such that $P_\Omega(x^0 - \nabla f(x^0)) \neq x^0$. In (Birgin *et al.* 2003b), at the expense of an extra gradient evaluation, it was used

$$\lambda_0 = \begin{cases} \max\{\lambda_{\min}, \min\{\bar{s}^\top \bar{s}/\bar{s}^\top \bar{y}, \lambda_{\max}\}\}, & \text{if } \bar{s}^\top \bar{y} > 0, \\ \lambda_{\max}, & \text{otherwise,} \end{cases}$$

where $\bar{s} = \bar{x} - x^0$, $\bar{y} = \nabla f(\bar{x}) - \nabla f(x^0)$, $\bar{x} = x^0 - \alpha_{\rm small} \nabla f(x^0)$,

$$\alpha_{\rm small} = \max\{\varepsilon_{\rm rel} \|x^0\|_\infty, \varepsilon_{\rm abs}\},$$

and $\varepsilon_{\rm rel}$ and $\varepsilon_{\rm abs}$ are relative and absolute small values related to the machine precision, respectively.

It is easy to see that SPG only needs $3n + O(1)$ double precision positions but one additional vector may be used to store the iterate with best functional value obtained by the method. This is almost always the last iterate but exceptions are possible. C/C++, Fortran 77 (including an interface with the **CUTEr** (Gould, Orban, and Toint 2003) test set), Matlab, and Octave subroutines implementing the Spectral Projected Gradient method (Algorithms 2.1 and 2.2) are available in the TANGO Project web page (`http://www.ime.usp.br/~egbirgin/tango/`). A Fortran 77 implementation is also available as Algorithm 813 of the ACM Collected Algorithms (`http://calgo.acm.org/`). An implementation of the SPG method in R (R Core Team 2012) is also available within the **BB** package Varadhan and Gilbert (2009). Recall that the method's purpose is to seek the least value of a function f of potentially many variables, subject to a convex set Ω onto which we know how to project. The objective function and its gradient should be coded by the user. Similarly, the user must provide a subroutine that computes the projection of an arbitrary vector x onto the feasible convex set Ω.

3. A matrix problem on the set of positive definite matrices

Optimization problems on the space of matrices, which are restricted to the convex set of

positive definite matrices, arise in various applications, such as statistics, as well as financial mathematics, model updating, and in general in matrix least-squares settings; see, e.g., (Boyd and Xiao 2005; Escalante and Raydan 1996; Fletcher 1985; Higham 2002; Hu and Olkin 1991; Yuan 2012).

To illustrate the use of the SPG method, we now describe a classification scheme (see, e.g., (Kawalec and Magdziak 2012)) that can be written as an optimization problem on the convex set of positive definite matrices. Given a training set of labelled examples

$$D = \{(z_i, w_i), i = 1, \ldots, m, z_i \in \mathbb{R}^q \text{ and } w_i \in \{1, -1\}\},$$

we aim to find a classifier ellipsoid $E(A, b) = \{y \in \mathbb{R}^q \mid y^\top A y + b^\top y = 1\}$ in $\mathbb{R}^q$ such that $z_i^\top A z_i + b^\top z_i \leq 1$ if $w_i = 1$ and $z_i^\top A z_i + b^\top z_i \geq 1$, otherwise. Since such ellipsoid may not exist, defining $I = \{i \in \{1, \ldots, m\} \mid w_i = 1\}$ and $O = \{i \in \{1, \ldots, m\} \mid w_i = -1\}$, we seek to minimize the function given by

$$f(A, b) = \frac{1}{m} \left[\sum_{i \in I} \max\{0, z_i^\top A z_i + b^\top z_i - 1\}^2 + \sum_{i \in O} \max\{0, 1 - z_i^\top A z_i - b^\top z_i\}^2 \right]$$

subject to symmetry and positive definiteness of the $q \times q$ matrix A. Moreover, to impose closed constraints on the semi-axes of the ellipsoid, we define a lower and an upper bound

$$0 < \hat{\lambda}_{\min} \leq \lambda_i(A) \leq \hat{\lambda}_{\max} < +\infty$$

on the eigenvalues $\lambda_i(A)$, $1 \leq i \leq q$. Given a square matrix A, its projection onto this closed and convex set can be computed in two steps (Escalante and Raydan 1996; Higham 1988). In the first step one symmetrizes A and in the second step one computes the $QDQ^\top$ decomposition of the symmetrized matrix and replaces its eigenvalues by their projection onto the interval $[\hat{\lambda}_{\min}, \hat{\lambda}_{\max}]$.

The number of variables n of the problem is $n = q(q + 1)$, corresponding to the q^2 elements of matrix $A \in \mathbb{R}^{q \times q}$ plus the elements of $b \in \mathbb{R}^q$. Variables correspond to the column wise representation of A plus the elements of b.

3.1. Coding the problem

The problem will be solved using the implementation of the SPG method that is part of the R package **BB** Varadhan and Gilbert (2009) as subroutine `spg`. Using the SPG method requires subroutines to compute the objective function, its gradient, and the projection of an arbitrary point onto the convex set. The gradient is computed once per iteration, while the objective function may be computed more than once. However, the nonmonotone strategy of the SPG method implies that the number of function evaluations per iteration is in general near one. There is a practical effect associated with this property. If the objective function and its gradient share common expressions, they may be computed jointly in order to save computational effort. This is the case of the problem at hand. Therefore, subroutines to compute the objective function and its gradient, that will be named `evalf` and `evalg`, respectively, are based on a subroutine named `evalfg` that, as its name suggests, computes both at the same time. Every time `evalf` is called, it calls `evalfg`, returns the computed value of the objective function and saves its gradient. When subroutine `evalg` is called, it returns the saved gradient. This is possible because the SPG method possesses the following

property: every time the gradient at a point x is required, the objective function was already evaluated at point x and that point was the last one at which the objective function was evaluated. Therefore, the code of subroutines that compute the objective function and its gradient is given by

```
R> evalf <- function(x) { r <- evalfg(x); gsaved <<- r$g; r$f }
```

and

```
R> evalg <- function(x) { gsaved }
```

Both subroutines return values effectively computed by the subroutine `evalfg` given by

```
R> evalfg <- function(x) {
+  n <- length(x)
+  A <- matrix(x,nrow=nmat,ncol=nmat)
+  b <- x[seq(length=nmat, from=nmat^2+1, by=1)]
+  fparc <- apply(p,2,function(y) {return( t(y) %*% A %*% y + b %*% y )}) - 1.0
+  I <- which( ( t=='i' & fparc>0.0 ) | ( t=='o' & fparc<0.0 ) )
+  if ( length(I) > 0 ) {
+      f <- sum( as.array(fparc[I])^2 ) / np
+      g <- ( as.vector( apply(as.matrix(p[,I]),2,
+           function(y) {return( c(as.vector(y %*% t(y)), y) )}) %*%
+           ( 2.0 * as.array(fparc[I]) ) ) / np
+      }
+  else {
+      f <- 0.0
+      g <- rep(0.0,n)
+      }
+
+  list(f=f,g=g)
+  }
```

The subroutine that computes the projection onto the feasible convex set is named `proj` and is given by

```
R> proj <- function(x) {
+  A <- matrix(x,nrow=nmat,ncol=nmat)
+  A <- 0.5 * ( t(A) + A )
+  r <- eigen(A,symmetric=TRUE)
+  lambda <- r$values
+  V <- r$vectors
+  lambda <- pmax( lowerl, pmin( lambda, upperl ) )
+  A <- V %*% diag(lambda) %*% t(V)
+  x[seq(length=nmat^2, from=1, by=1)] <- as.vector(A)
+  x
+  }
```

Note that subroutine `proj` uses subroutine `eigen` to compute eigenvalues and eigenvectors.

3.2. Numerical experiments

Numerical experiments were conducted using R version 2.14.1 on a 2.67GHz Intel Xeon CPU X5650 with 8GB of RAM memory and running GNU/Linux operating system (Ubuntu 12.04 LTS, kernel 3.2.0-33).

In the numerical experiments, we considered the default parameters of the SPG mentioned in Section 2. Regarding the initial spectral step length, we considered the choice given by (8) and for the stopping criterion we arbitrarily fixed $\varepsilon = 10^{-6}$. The nonmonotone line search parameter was arbitrarily set as $M = 100$.

Four small two-dimensional illustrative examples were designed. In the four examples we considered $q = 2$ and $m = 10,000$ random points $z_i \in \mathbb{R}^q$ uniformly distributed within the box $[-100, 100]^q$; and we fixed the suitable values $\hat{\lambda}_{\min} = 10^{-4}$ and $\hat{\lambda}_{\max} = 10^4$. In the first example, label $w_i = 1$ is given to the points inside a circle with radius 70 centered at the origin (see Figure 1(a)). If the reader is able to see the figure in colors, points labelled with $w_i = 1$ (i.e., inside the circle) are depicted in blue while points outside the circle (with $w_i = -1$) are depicted in red. Examples in Figures 1(b–d) correspond to a square of side 70 centered at the origin, a rectangle with height equal to 70 and width equal to 140 centered at the origin, and a triangle with corners $(-70, 0)$, $(0, -70)$, and $(70, 70)$, respectively.

The sequence of operations needed to solve the first example is given by:

```
R> library(BB)
R> nmat <- 2
R> np <- 10000
R> set.seed(123456)
R> p <- matrix(runif(nmat*np,min=-100.0,max=100.0),nrow=nmat,ncol=np)
R> t <- apply(p,2,function(y)
+       {if (sum(y^2)<=70.0^2) return('i') else return('o')})
R> lowerl <- rep(1.0e-04, nmat )
R> upperl <- rep(1.0e+04, nmat )
R> n <- nmat^2 + nmat
R> set.seed(123456)
R> x <- runif(n,min=-1.0,max=1.0)
R> ans <- spg(par=x, fn=evalf, gr=evalg, project=proj, method=1,
+             control=list(M=100, maxit=10000, maxfeval=100000,
+             ftol=0.0, gtol=1.e-06))
```

In the code above, `nmat` represents the dimension q of the space, `np` is the number of points m, and `p` saves the m randomly generated points $z_i \in \mathbb{R}^q$. For each point, `t` saves the value of w_i that, instead of -1 and 1, is represented by `'i'` that means inside and `'o'` that means outside the circle centered at the origin with radius 70. `lowerl` and `upperl` represent, respectively, the lower and the upper bounds $\hat{\lambda}_{\min}$ and $\hat{\lambda}_{\max}$ for the eigenvalues of matrix A. Finally, the number of variables `n` of the optimization problem and a random initial guess `x` $\in [-1, 1]^n$ are computed and the `spg` subroutine is called. The other three examples can be solved substituting the command line that defines `t` by

```
R> t <- apply(p,2,function(y) {if (length(y)==sum(-70.0<=y & y<=70.0))
+       return('i') else return('o')})

R> t <- apply(p,2,function(y) {if (-70.0<=y[1] & y[1]<=70.0 & -35.0<=y[2] &
+       y[2]<=35.0) return('i') else return('o')})
```

or

```
R> t <- apply(p,2,function(y) {if (2.0 * y[1] - y[2] <= 70.0 &
+       -y[1] + 2.0 * y[2] <= 70.0 &-y[1] - y[2] <= 70.0)
+       return('i') else return('o')})
```

respectively.

Figure 1 shows the solutions to the four illustrative examples. The problem depicted on Figure 1(a) corresponds to a problem with known global solution at which the objective function vanishes, and the graphic shows that the global solution is correctly found by the method. The remaining three examples clearly correspond to problems at which the global minimum is strictly positive. Moreover, problems depicted on Figure 1(b) and 1(c) are symmetric problems whose solutions are given by ellipses with their axes parallel to the Cartesian axes, while Figure 1(d) displays a solution with a rotated ellipse.

Table 1 presents a few figures that represent the computational effort needed by the SPG method to solve the problems. Figures in the table show that the stopping criterion was satisfied in the four cases and that, as expected, the ratio between the number of functional evaluations and the number of iterations is near unity, stressing the high rate of acceptance of the trial spectral step length along the projected gradient direction. A short note regarding the reported CPU times is in order. A Fortran 77 version of the SPG method (available in http://www.ime.usp.br/~egbirgin/) was also considered to solve the same four problems. Equivalent solutions were obtained using CPU times three orders of magnitude smaller than the ones required by the R version of the method. From the required total CPU time, approximately 99% is used to compute the objective function and its gradient. This observation is in complete agreement with the fact that SPG iterations have a time complexity $O(n)$ while the objective function and gradient evaluations have time complexity $O(nm)$, and that, in the considered examples, we have $n = q(q+1) = 6$ and $m = 10,000$. In this scenario, coding the problem subroutines (that computes the objective function f, its gradient, and the projection onto the convex feasible set Ω) in R and developing an interface to call a compiled Fortran 77 version of `spg` is not an option. On the other hand, when the execution of the subroutines that define the problem is inexpensive compared to the linear algebra involved in an iteration of the optimization method, coding the problem subroutines in an interpreted language like R and developing interfaces to optimization methods developed in compiled languages might be an adequate choice. This combination may jointly deliver (a) the fast development supplied by the usage of the user preferred language to code the problem subroutines and (b) the efficiency of powerful optimization tools developed in compiled languages especially suited to numeric computation and scientific computing. This is the case of, for example, the optimization method **Algencan** (Andreani, Birgin, Martínez, and Schuverdt 2007, 2008) (see also the TANGO Project web page: http://www.ime.usp.br/~egbirgin/tango/) for nonlinear programming problems, coded in Fortran 77 and with interfaces to R, AMPL, C/C++, **CUTEr**, Java, MATLAB, Octave, Python, and TCL.

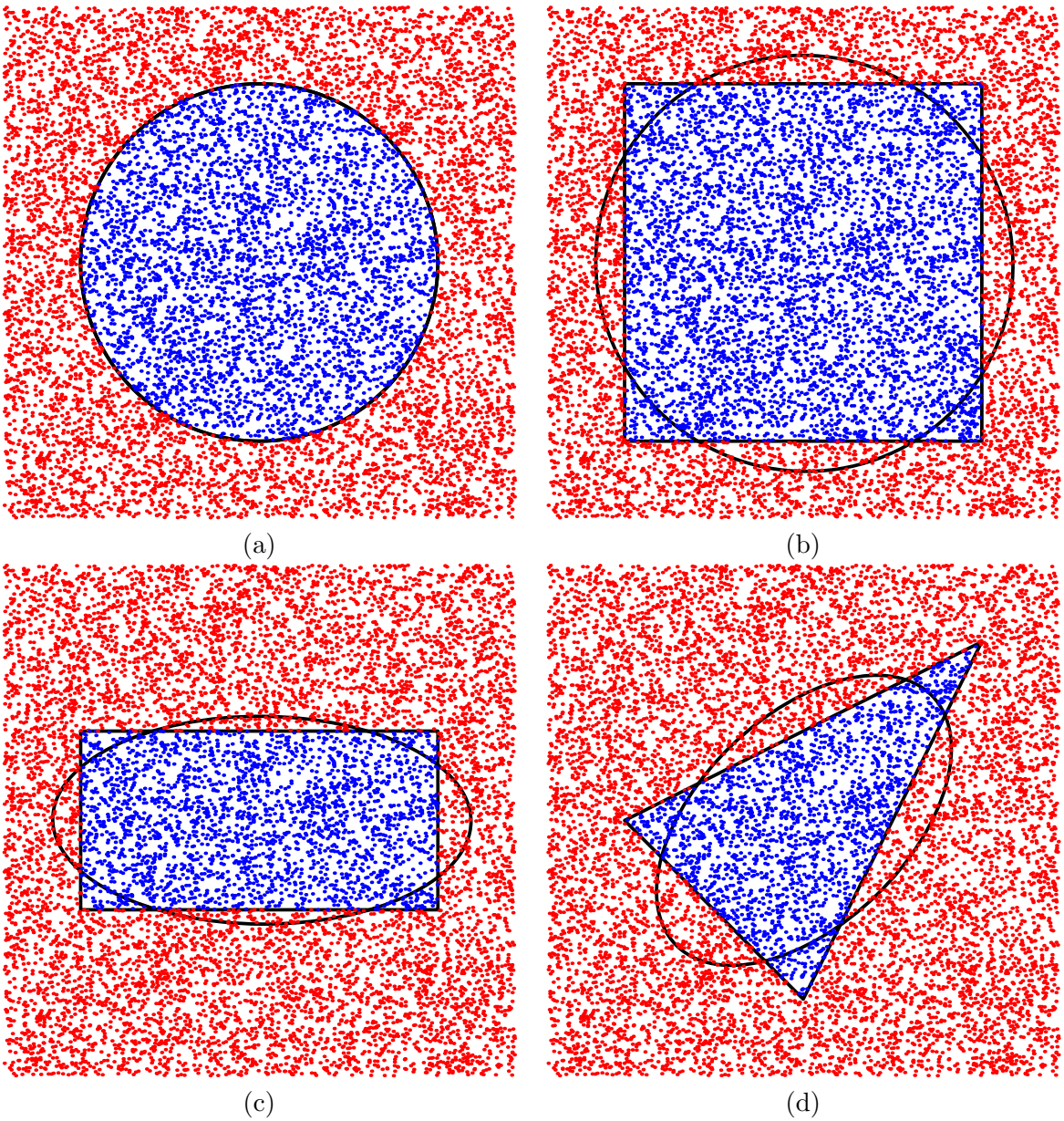

Figure 1: Solutions to four small two-dimensional illustrative examples.

Problem	# it	# fe	$f(x^*)$	$\|P_\Omega(x^* - \nabla f(x^*)) - x^*\|_\infty$	CPU Time
Circle	3307	3440	1.195192e−13	3.4e−07	1003.92
Square	1761	1907	2.352849e−03	7.9e−07	597.43
Rectangle	7269	8177	1.036716e−03	8.9e−07	2174.63
Triangle	7193	7753	6.512737e−03	9.9e−07	2199.36

Table 1: Performance of the R version of SPG in the four small two-dimensional illustrative examples.

4. Applications and extensions

As a general rule, the SPG method is applicable to large-scale convex constrained problems in which the projection onto the feasible set can be inexpensively computed. Moreover, the scenario is also beneficial to the application of the SPG method whenever the convex feasible set is hard to describe with nonlinear constraints or when the nonlinear constraints that are needed to describe the convex feasible set have an undesirable property (e.g., being too many, being highly nonlinear, or having a dense Jacobian). A clear example of this situation is the problem that illustrates the previous section. Comparisons with other methods for the particular case in which the feasible set is an n-dimensional box can be found in (Birgin *et al.* 2000; Birgin and Gentil 2012).

Since its appearance, the SPG method has been useful for solving real applications in different areas, including optics (Andrade, Birgin, Chambouleyron, Martínez, and Ventura 2008; Azofeifa, Clark, and Vargas 2005; Birgin, Chambouleyron, and Martínez 1999b; Birgin, Chambouleyron, Martínez, and Ventura 2003a; Chambouleyron, Ventura, Birgin, and Martínez 2009; Curiel, Vargas, and Barrera 2002; Mulato, Chambouleyron, Birgin, and Martínez 2000; Murphy 2007; Ramirez-Porras and Vargas-Castro 2004; Vargas 2002; Vargas, Azofeifa, and Clark 2003; Ventura, Birgin, Martínez, and Chambouleyron 2005), compressive sensing (van den Berg and Friedlander 2008, 2011; Figueiredo, Nowak, and Wright 2007; Loris, Bertero, Mol, Zanella, and Zanni 2009), geophysics (Bello and Raydan 2007; Birgin, Biloti, Tygel, and Santos 1999a; Cores and Loreto 2007; Deidda, Bonomi, and Manzi 2003; Zeev, Savasta, and Cores 2006), statistics (Borsdorf, Higham, and Raydan 2010; Varadhan and Gilbert 2009), image restoration (Benvenuto, Zanella, Zanni, and Bertero 2010; Bonettini, Zanella, and Zanni 2009; Bouhamidi, Jbilou, and Raydan 2011; Guerrero, Raydan, and Rojas 2012), atmospheric sciences (Jiang 2006; Mu, Duan, and Wang 2003), chemistry (Francisco, Martínez, and Martínez 2006; Birgin, Martínez, Martínez, and Rocha 2013), and dental radiology (Kolehmainen, Vanne, Siltanen, Järvenpää, Kaipio, Lassas, and Kalke 2006, 2007). The SPG method has also been combined with several schemes for solving scientific problems that appear in other areas of computational mathematics, including eigenvalue complementarity (Júdice, Raydan, Rosa, and Santos 2008), support vector machines (Cores, Escalante, Gonzalez-Lima, and Jiménez 2009; Dai and Fletcher 2006; Serafini, Zanghirati, and Zanni 2005), non-differentiable optimization (Crema, Loreto, and Raydan 2007), trust-region globalization (Maciel, Mendonça, and Verdiell 2012), generalized Sylvester equations (Bouhamidi *et al.* 2011), nonlinear monotone equations (Zhang and Zhou 2006), condition number estimation (Brás, Hager, and Júdice 2012), optimal control (Birgin and Evtushenko 1998), bound-constrained minimization (Andreani, Birgin, Martínez, and Schuverdt 2010; Andretta, Birgin, and Martínez 2005; Birgin and Martínez 2001, 2002), nonlinear programming (Andreani *et al.* 2007, 2008; Andreani, Birgin, Martínez, and Yuan 2005; Andretta, Birgin, and Martínez 2010; Birgin and Martínez 2008; Diniz-Ehrhardt, Gomes-Ruggiero, Martínez, and Santos 2004; Gomes-Ruggiero, Martínez, and Santos 2009), non-negative matrix factorization (Li, Liu, and Zheng 2012), and topology optimization (Tavakoli and Zhang 2012). Moreover, alternative choices of the spectral step length have been considered and analyzed for solving some related nonlinear problems. A spectral approach has been considered to solve large-scale nonlinear systems of equations using only the residual vector as search direction (Grippo and Sciandrone 2007; La Cruz and Raydan 2003; La Cruz *et al.* 2006; Varadhan and Gilbert 2009; Zhang and Zhou 2006). Spectral variations have also been developed for accelerating the convergence of fixed-point iterations, in connection with the well-known EM

algorithm which is frequently used in computational statistics (Roland and Varadhan 2005; Roland, Varadhan, and Frangakis 2007; Varadhan and Roland 2008).

The case in which the convex feasible set of the problem to be solved by SPG is defined by linear equality and inequality constraints has been considered in (Birgin *et al.* 2003b; Andreani *et al.* 2005; Martínez, Pilotta, and Raydan 2005; Andretta *et al.* 2010). A crucial observation is that this type of set is not necessarily one in which it is easy to project. Projecting onto a polytope requires to solve a convex quadratic programming problem, for which there exist many efficient algorithms, especially when the set exhibits additional structure. (To perform this task, the **BB** package relies on the method introduced in (Goldfarb and Idnani 1983), implemented in the package **quadprog** (Weingessel 2013).) An important observation is that the dual of such a problem is a concave box-constrained problem, which allows the use of specialized methods to solve them (see (Andreani *et al.* 2005)). In the large-scale case, it is important to have the possibility of performing the projection only aproximately. The SPG theory for that case have been extensively developed in (Birgin *et al.* 2003b).

5. Conclusions

The SPG method is nowadays a well-established numerical scheme for solving large-scale convex constrained optimization problems when the projection onto the feasible set can be performed efficiently. The attractiveness of the SPG method is mainly based on its simplicity. No sophisticated linear codes are required, no additional linear algebra packages are needed and the user can code its own version of the method with a relatively small effort. In this review, we presented the basic ideas behind the SPG method, discussed some of its most relevant properties, and discussed recent applications and extensions. In addition, we illustrated how the method can be applied to a particular matrix problem for which the feasible convex set is *easily* described by a subroutine that computes the projection onto it.

Acknowledgments

This work was supported by PRONEX-CNPq/FAPERJ (E-26/111.449/2010-APQ1), FAPESP (CEPID–Industrial Mathematics, grant 2013/07375-0), CNPq, and Fonacit-CNRS (Project No. 201200140).

References

Andrade R, Birgin EG, Chambouleyron I, Martínez JM, Ventura SD (2008). "Estimation of the Thickness and the Optical Parameters of Several Superimposed Thin Films Using Optimization." *Applied Optics*, **47**, 5208–5220.

Andreani R, Birgin EG, Martínez JM, Schuverdt ML (2007). "On Augmented Lagrangian Methods with General Lower-Level Constraints." *SIAM Journal on Optimization*, **18**, 1286–1309.

Andreani R, Birgin EG, Martínez JM, Schuverdt ML (2008). "Augmented Lagrangian Meth-

ods under the Constant Positive Linear Dependence Constraint Qualification." *Mathematical Programming*, **111**, 5–32.

Andreani R, Birgin EG, Martínez JM, Schuverdt ML (2010). "Second-Order Negative-Curvature Methods for Box-Constrained and General Constrained Optimization." *Computational Optimization and Applications*, **45**, 209–236.

Andreani R, Birgin EG, Martínez JM, Yuan JY (2005). "Spectral Projected Gradient and Variable Metric Methods for Optimization with Linear Inequalities." *IMA Journal of Numerical Analysis*, **25**, 221–252.

Andretta M, Birgin EG, Martínez JM (2005). "Practical Active-Set Euclidian Trust-Region Method with Spectral Projected Gradients for Bound-Constrained Minimization." *Optimization*, **54**, 305–325.

Andretta M, Birgin EG, Martínez JM (2010). "Partial Spectral Projected Gradient Method with Active-Set Strategy for Linearly Constrained Optimization." *Numerical Algorithms*, **53**(1), 23–52.

Asmundis RD, Serafino DD, Riccio F, Toraldo G (2012). "On Spectral Properties of Steepest Descent Methods." http://www.optimization-online.org/DB_HTML/2012/02/3349.html.

Azofeifa D, Clark N, Vargas W (2005). "Optical and Electrical Properties of Terbium Films as a Function of Hydrogen Concentration." *Physica Status Solidi B – Basic Solid State Physics*, **242**, 2005–2009.

Barzilai J, Borwein JM (1988). "Two Point Step Size Gradient Methods." *IMA Journal of Numerical Analysis*, **8**, 141–148.

Bello L, Raydan M (2007). "Convex Constrained Optimization for the Seismic Reflection Tomography Problem." *Journal of Applied Geophysics*, **62**, 158–166.

Benvenuto F, Zanella R, Zanni L, Bertero M (2010). "Nonnegative Least-Squares Image Deblurring: Improved Gradient Projection Approaches." *Inverse Problems*, **26**, 025004.

Berlinet AF, Roland C (2011). "Geometric Interpretation of Some Cauchy Related Methods." *Numerische Mathematik*, **119**, 437–464.

Bernaards CA, Jennrich RI (2005). "Gradient Projection Algorithms and Software for Arbitrary Rotation Criteria in Factor Analysis." *Educational and Psychological Measurement*, **65**, 676–696.

Bertsekas DP (1976). "On the Goldstein-Levitin-Polyak Gradient Projection Method." *IEEE Transactions on Automatic Control*, **21**, 174–184.

Bertsekas DP (1999). *Nonlinear Programming*. Athena Scientific, Boston.

Birgin EG, Biloti R, Tygel M, Santos LT (1999a). "Restricted Optimization: a Clue to a Fast and Accurate Implementation of the Common Reflection Surface Stack Method." *Journal of Applied Geophysics*, **42**, 143–155.

Birgin EG, Chambouleyron I, Martínez JM (1999b). "Estimation of the Optical Constants and the Thickness of Thin Films Using Unconstrained Optimization." *Journal of Computational Physics*, **151**, 862–880.

Birgin EG, Chambouleyron I, Martínez JM, Ventura SD (2003a). "Estimation of Optical Parameters of Very Thin Films." *Applied Numerical Mathematics*, **47**, 109–119.

Birgin EG, Evtushenko YG (1998). "Automatic Differentiation and Spectral Projected Gradient Methods for Optimal Control Problems." *Optimization Methods & Software*, **10**, 125–146.

Birgin EG, Gentil JM (2012). "Evaluating Bound-Constrained Minimization Software." *Computational Optimization and Applications*, **53**, 347–373.

Birgin EG, Martínez JM (2001). "A Box-Constrained Optimization Algorithm with Negative Curvature Directions and Spectral Projected Gradients." *Computing [Suppl]*, **15**, 49–60.

Birgin EG, Martínez JM (2002). "Large-Scale Active-Set Box-Constrained Optimization Method with Spectral Projected Gradients." *Computational Optimization and Applications*, **23**, 101–125.

Birgin EG, Martínez JM (2008). "Structured Minimal-Memory Inexact Quasi-Newton Method and Secant Preconditioners for Augmented Lagrangian Optimization." *Computational Optimization and Applications*, **39**, 1–16.

Birgin EG, Martínez JM, Martínez L, Rocha GB (2013). "Sparse projected-gradient method as a linear-scaling low-memory alternative to diagonalization in self-consistent field electronic structure calculations." *Journal of Chemical Theory and Computation*, **9**, 1043–1051.

Birgin EG, Martínez JM, Raydan M (2000). "Nonmonotone Spectral Projected Gradient Methods on Convex Sets." *SIAM Journal on Optimization*, **10**, 1196–1211.

Birgin EG, Martínez JM, Raydan M (2001). "Algorithm 813: SPG - Software for Convex-Constrained Optimization." *ACM Transactions on Mathematical Software*, **27**, 340–349.

Birgin EG, Martínez JM, Raydan M (2003b). "Inexact Spectral Projected Gradient Methods on Convex Sets." *IMA Journal of Numerical Analysis*, **23**, 539–559.

Bonettini S, Zanella R, Zanni L (2009). "A Scaled Gradient Projection Method for Constrained Image Deblurring." *Inverse Problems*, **25**, 015002.

Borsdorf R, Higham NJ, Raydan M (2010). "Computing a Nearest Correlation Matrix with Factor Structure." *SIAM Journal on Matrix Analysis and Applications*, **31**, 2603–2622.

Bouhamidi A, Jbilou K, Raydan M (2011). "Convex Constrained Optimization for Large-Scale Generalized Sylvester Equations." *Computational Optimization and Applications*, **48**, 233–253.

Boyd S, Xiao L (2005). "Least-Squares Covariance Matrix Adjustment." *SIAM Journal on Matrix Analysis and Applications*, **27**, 532–546.

Brás CP, Hager WW, Júdice JJ (2012). "An Investigation of Feasible Descent Algorithms for Estimating the Condition Number of a Matrix." *TOP - Journal of the Spanish Society of Statistics and Operations Research*, **20**, 791–809.

Cauchy A (1847). "Méthode Générale pour la Résolution des Systèmes D'équations Simultanées." *Compte-Rendu de l'Académie des Sciences*, **27**, 536–538.

Chambouleyron I, Ventura SD, Birgin EG, Martínez JM (2009). "Optical Constants and Thickness Determination of Very Thin Amorphous Semiconductor Films." *Journal of Applied Physics*, **92**, 3093–3102.

Cores D, Escalante R, Gonzalez-Lima M, Jiménez O (2009). "On the Use of the Spectral Projected Gradient Method for Support Vector Machines." *Computational and Applied Mathematics*, **28**, 327–364.

Cores D, Loreto M (2007). "A Generalized Two-Point Ellipsoidal Anisotropic Ray Tracing for Converted Waves." *Optimization and Engineering*, **8**, 373–396.

Crema A, Loreto M, Raydan M (2007). "Spectral Projected Subgradient with a Momentum Term for the Lagrangean Dual Approach." *Computers and Operations Research*, **34**, 3174–3186.

Curiel F, Vargas WE, Barrera RG (2002). "Visible Spectral Dependence of the Scattering and Absorption Coefficients of Pigmented Coatings from Inversion of Diffuse Reflectance Spectra." *Applied Optics*, **41**, 5969–5978.

Dai YH, Fletcher R (2005). "On the Asymptotic Behaviour of Some New Gradient Methods." *Mathematical Programming*, **103**, 541–559.

Dai YH, Fletcher R (2006). "New Algorithms for Singly Linearly Constrained Quadratic Programs Subject to Lower and Upper Bounds." *Mathematical Programming*, **106**, 403–421.

Dai YH, Liao LZ (2002). "R-Linear Convergence of the Barzilai and Borwein Gradient Method." *IMA Journal on Numerical Analysis*, **22**, 1–10.

Dai YH, Zhang HC (2001). "Adaptive Two-Point Stepsize Gradient Algorithm." *Numerical Algorithms*, **27**, 377–385.

Deidda GP, Bonomi E, Manzi C (2003). "Inversion of Electrical Conductivity Data With Tikhonov Regularization Approach: Some Considerations." *Annals of Geophysics*, **46**, 549–558.

Dennis JE, Moré JJ (1977). "Quasi-Newton Methods, Motivation and Theory." *SIAM Review*, **19**, 46–89.

Dennis JE, Schnabel RB (1983). *Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, NJ.

Diniz-Ehrhardt MA, Gomes-Ruggiero MA, Martínez JM, Santos SA (2004). "Augmented Lagrangian Algorithms Based on the Spectral Projected Gradient Method for Solving Nonlinear Programming Problems." *Journal of Optimization Theory and Applications*, **123**, 497–517.

Escalante R, Raydan M (1996). "Dykstra's Algorithm for a Constrained Least-Squares Matrix Problem." *Numerical Linear Algebra and Applications*, **3**, 459–471.

Figueiredo MA, Nowak RD, Wright SJ (2007). "Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems." *IEEE Journal of Selected Topics in Signal Processing*, **1**, 586–597.

Fletcher R (1985). "Semi-Definite Matrix Constraints in Optimization." *SIAM Journal on Control and Optimization*, **23**, 493–513.

Fletcher R (1990). "Low Storage Methods for Unconstrained Optimization." *Lectures in Applied Mathematics*, **26**, 165–179.

Fletcher R (2005). "On the Barzilai-Borwein Method." In L Qi and K Teo and X Yang (ed.), *Optimization and Control with Applications*, volume of *Series in Applied Optimization, Volume 96*, pp. 235–256. Kluwer.

Francisco JB, Martínez JM, Martínez L (2006). "Density-Based Globally Convergent Trust-Region Methods for Self-Consistent Field Electronic Structure Calculations." *Journal of Mathematical Chemistry*, **40**, 349–377.

Friedlander A, Martínez JM, Molina B, Raydan M (1999). "Gradient Method with Retards and Generalizations." *SIAM Journal on Numerical Analysis*, **36**, 275–289.

Glunt W, Hayden TL, Raydan M (1993). "Molecular Conformations from Distance Matrices." *Journal of Computational Chemistry*, **14**, 114–120.

Goldfarb D, Idnani A (1983). "A numerically stable dual method for solving strictly convex quadratic programs." *Mathematical Programming*, **27**, 1–33.

Goldstein AA (1964). "Convex Programming in Hilbert Space." *Bulletin of the American Mathematical Society*, **70**, 709–710.

Gomes-Ruggiero MA, Martínez JM, Santos SA (2009). "Spectral Projected Gradient Method with Inexact Restoration for Minimization with Nonconvex Constraints." *SIAM Journal on Scientific Computing*, **31**, 1628–1652.

Gould NIM, Orban D, Toint PL (2003). "CUTEr and SifDec: A Constrained and Unconstrained Testing Environment, Revisited." *ACM Transactions on Mathematical Software*, **29**, 373–394.

Grippo L, Lampariello F, Lucidi S (1986). "A Nonmonotone Line Search Technique for Newton's Method." *SIAM Journal on Numerical Analysis*, **23**, 707–716.

Grippo L, Lampariello F, Lucidi S (1991). "A Class of Nonmonotone Stabilization Methods in Unconstrained Optimization." *Numerische Mathematik*, **59**, 779–805.

Grippo L, Sciandrone M (2002). "Nonmonotone Globalization Techniques for the Barzilai-Borwein Gradient Method." *Computational Optimization and Application*, **23**, 143–169.

Grippo L, Sciandrone M (2007). "Nonmonotone Derivative-Free Methods for Nonlinear Equations." *Computational Optimization and Application*, **37**, 297–328.

Guerrero J, Raydan M, Rojas M (2012). "A Hybrid-Optimization Method for Large-Scale Non-Negative Full Regularization in Image Restoration." *Inverse Problems in Science & Engineering*. doi:10.1080/17415977.2012.720684.

Hestenes MR, Stiefel E (1952). "Methods of Conjugate Gradients for Solving Linear systems." *Journal of Research of the National Bureau of Standards*, **49**, 409–436.

Higham NJ (1988). "Computing a Nearest Symmetric Positive Semidefinite Matrix." *Linear Algebra and its Applications*, **103**, 103–118.

Higham NJ (2002). "Computing the Nearest Correlation Matrix – A Problem from Finance." *IMA Journal of Numerical Analysis*, **22**, 329–343.

Hu H, Olkin I (1991). "A Numerical Procedure for Finding the Positive Definite Matrix Closest to a Patterned Matrix." *Statistics and Probability letters*, **12**, 511–515.

Jiang Z (2006). "Applications of Conditional Nonlinear Optimal Perturbation to the Study of the Stability and Sensitivity of the Jovian Atmosphere." *Advances in Atmospheric Sciences*, **23**, 775–783.

Júdice JJ, Raydan M, Rosa SS, Santos SA (2008). "On the Solution of the Symmetric Eigenvalue Complementarity Problem by the Spectral Projected Gradient Algorithm." *Numerical Algorithms*, **47**, 391–407.

Kawalec A, Magdziak M (2012). "Usability Assessment of Selected Methods of Optimization for Some Measurement Task in Coordinate Measurement Technique." *Measurement*, **45**, 2330–2338.

Kolehmainen V, Vanne A, Siltanen S, Järvenpää S, Kaipio JP, Lassas M, Kalke M (2006). "Parallelized Bayesian Inversion for Three-Dimensional Dental X-ray Imaging." *IEEE Transactions on Medical Imaging*, **25**, 218–228.

Kolehmainen V, Vanne A, Siltanen S, Järvenpää S, Kaipio JP, Lassas M, Kalke M (2007). "Bayesian Inversion Method for 3D Dental X-ray Imaging." *e & i Elektrotechnik und Informationstechnik*, **124**, 248–253.

La Cruz W, Martínez JM, Raydan M (2006). "Spectral Residual Method Without Gradient Information for Solving Large-Scale Nonlinear Systems of Equations." *Mathematics of Computation*, **75**, 1429–1448.

La Cruz W, Raydan M (2003). "Nonmonotone Spectral Methods for Large-Scale Nonlinear Systems." *Optimization Methods and Software*, **18**, 583–599.

Levitin ES, Polyak BT (1966). "Constrained Minimization Problems." *USSR Computational Mathematics and Mathematical Physics*, **6**, 1–50.

Li X, Liu H, Zheng X (2012). "Non-Monotone Projection Gradient Method for Non-Negative Matrix Factorization." *Computational Optimization and Applications*, **51**, 1163–1171.

Loris I, Bertero M, Mol CD, Zanella R, Zanni L (2009). "Accelerating Gradient Projection Methods for ℓ_1-Constrained Signal Recovery by Steplength Selection Rules." *Applied and Computational Harmonic Analysis*, **27**, 247–254.

Maciel MC, Mendonça MG, Verdiell AB (2012). "Monotone and Nonmonotone Trust-Region-Based Algorithms for Large Scale Unconstrained Optimization Problems." *Computational Optimization and Applications*. doi:10.1007/s10589-012-9477-8.

Martínez JM, Pilotta EA, Raydan M (2005). "Spectral gradient methods for linearly constrained optimization." *Journal of Optimization Theory and Applications*, **125**, 629–651.

Molina B, Raydan M (1996). "Preconditioned Barzilai-Borwein Method for the Numerical Solution of Partial Differential Equations." *Numerical Algorithms*, **13**, 45–60.

Mu M, Duan WS, Wang B (2003). "Conditional Nonlinear Optimal Perturbation and its Applications." *Nonlinear Processes in Geophysics*, **10**, 493–501.

Mulato M, Chambouleyron I, Birgin EG, Martínez JM (2000). "Determination of Thickness and Optical Constants of a-Si:H Films from Transmittance Data." *Applied Physics Letters*, **77**, 2133–2135.

Murphy AB (2007). "Optical Properties of an Optically Rough Coating from Inversion of Diffuse Reflectance Measurements." *Applied Optics*, **46**, 3133–3143.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Ramirez-Porras A, Vargas-Castro WE (2004). "Transmission of Visible Light Through Oxidized Copper Films: Feasibility of Using a Spectral Projected Gradient Method." *Applied Optics*, **43**, 1508–1514.

Raydan M (1993). "On the Barzilai and Borwein Choice of Steplength for the Gradient Method." *IMA Journal of Numerical Analysis*, **13**, 321–326.

Raydan M (1997). "The Barzilai and Borwein Gradient Method for the Large Scale Unconstrained Minimization Problem." *SIAM Journal on Optimization*, **7**, 26–33.

Raydan M, Svaiter BF (2002). "Relaxed Steepest Descent and Cauchy-Barzilai-Borwein Method." *Computational Optimization and Applications*, **21**, 155–167.

Roland C, Varadhan R (2005). "New Iterative Schemes for Nonlinear Fixed Point Problems, with Applications to Problems with Bifurcations and Incomplete-Data Problems." *Applied Numerical Mathematics*, **55**, 215–226.

Roland C, Varadhan R, Frangakis CE (2007). "Squared Polynomial Extrapolation Methods with Cycling: An Application to the Positron Emission Tomography Problem." *Numerical Algorithms*, **44**, 159–172.

Serafini T, Zanghirati G, Zanni L (2005). "Gradient Projection Methods for Quadratic Programs and Applications in Training Support Vector Machines." *Optimization Methods and Software*, **20**, 347–372.

Tavakoli R, Zhang H (2012). "A Nonmonotone Spectral Projected Gradient Method for Large-Scale Topology Optimization Problems." *Numerical Algebra, Control and Optimization*, **2**, 395–412.

Toint PL (1996). "An Assessment of Nonmonotone Linesearch Techniques for Unconstrained Optimization." *SIAM Journal on Scientific Computing*, **17**, 725–739.

van den Berg E, Friedlander MP (2008). "Probing the Pareto frontier for basis pursuit solutions." *SIAM Journal on Scientific Computing*, **31**, 890–912.

van den Berg E, Friedlander MP (2011). "Sparse Optimization with Least-Squares Constraints." *SIAM Journal on Optimization*, **21**, 1201–1229.

van den Doel K, Ascher U (2012). "The Chaotic Nature of Faster Gradient Descent Methods." *Journal of Scientific Computing*, **51**, 560–581.

Varadhan R, Gilbert PD (2009). "BB: An R Package for Solving a Large System of Nonlinear Equations and for Optimizing a High-Dimensional Nonlinear Objective Function." *Journal of Statistical Software*, **32**, 1–26.

Varadhan R, Roland C (2008). "Simple, Stable, and General Methods for Accelerating Any EM Algorithm." *Scandinavian Journal of Statistics*, **35**, 335–353.

Vargas WE (2002). "Inversion Methods from Kiabelka-Munk Analysis." *Journal of Optics A - Pure and Applied Optics*, **4**, 452–456.

Vargas WE, Azofeifa DE, Clark N (2003). "Retrieved Optical Properties of Thin Films on Absorbing Substrates from Transmittance Measurements by Application of a Spectral Projected Gradient Method." *Thin Solid Films*, **425**, 1–8.

Ventura SD, Birgin EG, Martínez JM, Chambouleyron I (2005). "Optimization Techniques for the Estimation of the Thickness and the Optical Parameters of Thin Films Using Reflectance Data." *Journal of Applied Physics*, **97**, 043512.

Weingessel A (2013). "**quadprog**: Functions to solve Quadratic Programming Problems. R package version 1.5-5." URL `http://cran.r-project.org/package=quadprog`.

Yuan Q (2012). "Matrix Linear Variational Inequality Approach for Finite Element Model Updating." *Mechanical Systems and Signal Processing*, **28**, 507–516.

Zeev N, Savasta O, Cores D (2006). "Non-Monotone Spectral Projected Gradient Method Applied to Full Waveform Inversion." *Geophysical Prospecting*, **54**, 525–534.

Zhang H, Hager WW (2004). "A Nonmonotone Line Search Technique and its Application to Unconstrained Optimization." *SIAM Journal on Optimization*, **14**, 1043–1056.

Zhang L, Zhou W (2006). "Spectral Gradient Projection Method for Solving Nonlinear Monotone Equations." *Journal of Computational and Applied Mathematics*, **196**, 478–484.

Affiliation:

E. G. Birgin
Department of Computer Science
Institute of Mathematics and Statistics
University of São Paulo 05508-090, São Paulo, SP, Brazil
e-mail: egbirgin@ime.usp.br

J. M. Martínez
Department of Applied Mathematics
Institute of Mathematics, Statistics and Scientific Computing
University of Campinas, Campinas, SP, Brazil
e-mail: martinez@ime.unicamp.br

M. Raydan
Departamento de Cómputo Científico y Estadística
Universidad Simón Bolívar
Ap. 89000, Caracas 1080-A, Venezuela
e-mail: mraydan@usb.ve

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume VV, Issue II *Submitted:* yyyy-mm-dd
MMMMMM YYYY *Accepted:* yyyy-mm-dd