Universidad Central de Venezuela
Facultad de Ciencias
Escuela de Computación

Lecturas en Ciencias de la Computación
ISSN 1316-6239

ON THE CONVERGENCE OF SPECTRAL PROJECTED SUBGRADIENT FOR THE LAGRANGEAN DUAL APPROACH

Milagros Loreto y Alejandro Crema

RT 2008-04

Centro de Cálculo Científico y Tecnológico de la UCV
CCCT-UCV
Centro de Investigación de Operaciones y Modelos Matemáticos de la UCV
CIOMMA-UCV
Caracas, Junio, 2008.

ON THE CONVERGENCE OF SPECTRAL PROJECTED SUBGRADIENT FOR THE LAGRANGEAN DUAL APPROACH

Milagros Loreto * Alejandro Crema †

June 26, 2008

Abstract

A study of the convergence properties of spectral projected subgradient method is presented and the convergence is shown. The convergence is based on spectral projected gradient approach. Some updates of the spectral projected subgradient are described.

Key words: Spectral projected gradient, subgradient optimization, global convergence.

1 Introduction

The spectral projected gradient method is related to the practical version of Bertsekas [2] of the classical gradient projected method of Goldstein [13] and Levitin [16]. However, some crucial differences make this method much more efficient than its gradient projection predecessors. The key issue is that the first trial step at each iteration is taken using the spectral steplength introduced in [1] and later analyzed in [9], [10], [17] among others. The spectral

*Departamento de Computación, Facultad de Ciencias, Universidad Central de Venezuela, Ap. 47002, Caracas 1041-A, Venezuela (mloreto@kuaimare.ciens.ucv.ve). Supported by the Center of Scientific Computing at UCV.

†Departamento de Computación, Facultad de Ciencias, Universidad Central de Venezuela, Ap. 47002, Caracas 1041-A, Venezuela (acrema@kuaimare.ciens.ucv.ve).

step is a Rayleigh quotient related with an average Hessian matrix. For a review containing more recent advances on this special choice of steplength see [11].

Therefore, it is natural to transport the projected spectral gradient idea with a nonmonotone line search to the projected subgradient, in order to speed up the convergence of the subgradient method and to use the steplength that does not depend on the optimal value of the objective function. It just has been done by Crema *et al.*[7] so the subgradient method was embeded into a globalization strategy that accepts the spectral step as frequently as possible.

In this work we extend the spectral projected subgradient method recently proposed by Crema *et al.*[7]. We update the spectral projected subgradient with the version of spectral projected gradient calls SPG1 ([4, 5]) and we obtain a new version of the spectral projected subgradient calls $SPS2$. This new version allows to show convergence of spectral projected subgradient methods.

This paper is organized as follows. In section 2 we present the spectral projected subgradient algorithm with the updates coming from spectral projected gradient. We discuss the globalization strategy, which is based on non-monotone line search technique of Grippo, Lampariello and Lucidi [14] combined with the globalization scheme recently proposed by La Cruz *et al.* [8] along with some conditions for the spectral step. In section 3 we prove global convergence results based on globalization strategy and under some mild assumptions on the spectral step. Some final remarks are presented in section 4.

2 Spectral Projected Subgradient Algorithm

Let us consider the following integer programming problem (P)

$$\begin{array}{lc} \max & c^T x \\ \text{s.t.} & Ax \leq b \\ & Dx \leq e \\ & x \in \mathbb{Z}^n,\ x \geq 0 \end{array}$$

where $\mathbb{Z}$ represents the integer numbers, c, b and e are vectors, A and D are matrices of suitable dimensions.

We are mainly concerned with the Lagrangean dual formulation of P, that will be referred as problem (D) and is given by

$$\begin{array}{ll} \min & f(\lambda) \\ \text{s.t.} & \lambda \geq 0, \end{array}$$

where $f(\lambda) = \max\{c^T x + \lambda^T(b - Ax), x \in X\}$ is convex, piece-wise linear, and non differentiable at some points.
In this work we assume that X is bounded and hence finite and $\Omega = \{\lambda : \lambda \geq 0\}$ is nonempty, closed, and convex subset of $\mathbb{R}^n$.

This problem can be solved using spectral projected subgradiente [7] and subgradient algorithm [15] among others. A brief review of the ideas associated with duality for solving integer programming problems can be found in [3, 12, 15].

Given $\lambda_0 \in \Omega$, we define $P_\Omega(\lambda_0)$ as the projection of λ_0 on Ω. We denote g_k as the subgradient of f in λ_k. The algorithm starts with $\lambda_0 \in \Omega$ and uses an integer $M \geq 1$, a small parameter $\alpha_{min} > 0$, a large parameter $\alpha_{max} > \alpha_{min}$, a sufficient decrease parameter $\gamma \in (0,1)$ and safeguarding parameters $0 < \sigma_1 < \sigma_2 < 1$. Initially $\alpha_0 \in [\alpha_{min}, \alpha_{max}]$ is arbitrary, $\eta_0 = max(f(\lambda_0), \| g(\lambda_0) \|)$, $m_0 = 0$ and $\mu \in [0,1]$. Given $\lambda_k \in \Omega$ and $\alpha_k \in [\alpha_{min}, \alpha_{max}]$. The algorithm $SPS2$ is described, it obtains λ_{k+1} and α_{k+1} at iteration $k+1$.

ALGORITHM $SPS2$.

- Step 1.- Backtracking

    Step 1.1 Set $\tau = \alpha_k$, $m_k = \tau g_k + \mu m_{k-1}$

    Step 1.2 Set $\lambda_+ = P(\lambda_k - m_k)$, $\eta_k = \frac{\eta_0}{k^{1.1}}$

    Step 1.3

$$\text{While } f(\lambda_+) > \max_{0 \leq j \leq \min\{k, M-1\}} f(\lambda_{k-j}) + \gamma(\lambda_+ - \lambda_k)^t g_k + \eta_k$$

$$\text{Choose } \tau_{new} \in [\sigma_1 \tau, \sigma_2 \tau]$$

$$\tau = \tau_{new}$$

$$m_k = \tau g_k + \mu m_{k-1}$$

$$\lambda_+ = P(\lambda_k - m_k)$$

Step 1.4

$$\lambda_{k+1} = \lambda_+, \quad \tau_k = \tau_{new}, \quad s_k = \lambda_{k+1} - \lambda_k, \quad y_k = g_{k+1} - g_k$$

- Step 2.- Compute α_{k+1}

  Step 2.1 Compute $b_k = s_k^t y_k$

  Step 2.2 if $b_k \leq 0$, set $\alpha_{k+1} = \alpha_{\max}$ else, compute $a_k = s_k^t s_k$ and

$$\alpha_{k+1} = \min\{\alpha_{\max}, \max\{\alpha_{\min}, a_k/b_k\}\}$$

  Step 2.3

$$\text{if} \quad \alpha_{k+1} \geq \frac{10^8}{\log(k)} \quad \text{then} \quad \alpha_{k+1} = \frac{10^8}{\log(k)}$$

$$\text{if} \quad \alpha_{k+1} \leq \frac{10^{-8}}{\log(k)} \quad \text{then} \quad \alpha_{k+1} = \frac{10^{-8}}{\log(k)}$$

Remarks:

1) The steplength α_k is obtained using the spectral choice, the step 2.2 guarantees that $s_k^t y_k \neq 0$ and step 2.3 allows to bound $\{\alpha_k\}$ by one sequence that guarantees the following condition:

$$\alpha_k > 0 \quad \forall k, \quad \lim_{k\to\infty} \alpha_k = 0, \quad \sum_{k=1}^{\infty} \alpha_k = \infty, \tag{1}$$

We use the step 2.3 to bound $\{\alpha_k\}$ for one sequence that satisfies the condition (1).

2) The parameter τ_{new} could be computed in many different ways. As an example, Crema *et al.*[7] use $\tau_{new=\tau/2}$.

3) For our nonmonotone globalization technique, we combine and extend the Grippo, Lampariello and Lucidi [14] line search scheme with the globalization scheme recently proposed by La Cruz *et al.* [8], and used by Crema *et al.*[7]. Roughly speaking, our acceptance condition for the next iterate is

$$f(\lambda_{k+1}) \leq \max_{0\leq j\leq \min\{k, M-1\}} f(\lambda_{k-j}) + \gamma(\lambda_{k+1} - \lambda_k)^t g_k + \eta_k \tag{2}$$

where γ is a small positive number, and η_k is chosen such that

$$0 < \sum_k \eta_k < \infty. \tag{3}$$

The terms $\max_{0 \le j \le M-1} f(\lambda_{k-j})$ and $\eta_k > 0$ are responsible for the sufficiently nonmonotone behavior of $f(\lambda_k)$. The parameter $\eta_k = \eta_0 / k^{(1.1)}$ guarantees that (3) is satisfied. If we choose $\eta_k = \eta_0 / k^r$ where $r > 1$, then (3) will also hold. Our feasible choice $r = 1.1$ is suitable for the sufficiently nonmonotone desired behavior of the method.

4) If the parameter the momentum μ is zero the Step 1.2 changes to $\lambda_+ = P(\lambda_k - \tau g_k)$.

5) The $SPS2$ is stopped when MAXITER iterations are reached.

The $SPS2$ is similar to the algorithm SPS shown in [7] with a few differences such as: 1) λ_+ is projected in each iteration into the backtracking step. As a consequence, the scalar product $(\lambda_+ - \lambda_k)^t g_k$ in the nonmonotone globalization condition must be computed for each point λ_+. 2) The condition (1) is added to support theoretical convergence.

3 Convergence Results

We based the convergence analysis on the non-monotone line search technique ([14],[8]) and on some results of [6]. Since the spectral projected subgradient method is not a descent method, it is common to keep track of the best point found so far, *i.e.*, the one with smallest function value. At each step, we set

$$f_k^{best} = \min\{f_{k-1}^{best}, f(\lambda_k)\}$$

If λ_k is the best point found so far we have:

$$f_k^{best} = \min\{f(\lambda_1), \ldots, f(\lambda_{k-1}), f(\lambda_k)\},$$

i.e., f_k^{best} is the best objective value found in k iterations, as f_k^{best} is decreasing, it has a limit.

There are many results on convergence of the subgradient method. For constant step size and constant step length, the subgradient algorithm is guaranteed to converge within some range of the optimal value, we have:

$$\lim_{k\to\infty} f_k^{best} - f_* < \epsilon$$

where f_* denotes the optimal value of problem , *i.e.*, we kave $f^* = inf_\lambda f(\lambda)$. For the diminishing step size and step length rules, the algorithm is guaranteed to converge to the optimal value, *i.e.*, we have $\lim_{k\to\infty} f(\lambda_k) = f_*$.
For the following analysis of convergence, we suppose the momentum term is zero but it can be extended for momentum term different of zero.

3.1 Basics inequalities.

Inequality 3.1a:

For a nonmonotone line search technique by Grippo, Lampariello and Lucidi [14] we have:

$$f(\lambda_{k+1}) \leq \max_{0\leq j\leq m(k)} f(\lambda_{k-j}) + \gamma(\lambda_{k+1} - \lambda_k)^t g_k$$

where $0 \leq m(k) \leq \min[m(k-1)+1, M]$ and $m(0) = 0$, $k \geq 1$. Set $l(k)$ be an integer such that:

$$k - m(k) \leq l(k) \leq k,$$

$$f(\lambda_{l(k)}) = \max_{0\leq j\leq m(k)} f(\lambda_{k-j})$$

and the sequence $\{f(\lambda_{l(k)})\}$ is nonincreasing as shown in [14]. Now we observe that:

$$f_k^{best} < f(\lambda_{l(k)}) = \max_{0\leq j\leq m(k)} f(\lambda_{k-j}),$$

and we obtain the inequality:

$$f_k^{best} - f_* < \max_{0\leq j\leq m(k)} f(\lambda_{k-j}) - f_*. \tag{3.1a}$$

Inequality 3.1b:

Let $z_{k+1} = \lambda_k - \alpha_k g_k$ a standard spectral subgradient update before the projection on Ω, $\lambda_{k+1} = P_\Omega(z_{k+1})$ and λ_* an optimal solution. We have:

$$\| \lambda_{k+1} - \lambda_* \|_2 = \| P_\Omega(z_{k+1}) - \lambda_* \|_2 \leq \| z_{k+1} - \lambda_* \|_2$$

Using inequality above we have:

$$\| z_{k+1} - \lambda_* \|_2^2 = \| \lambda_k - \alpha_k g_k - \lambda_* \|_2^2 \leq \| \lambda_k - \lambda_* \|_2^2 - 2\alpha_k g_k^T(\lambda_k - \lambda_*) + \alpha_k^2 \| g_k \|_2^2$$

and we obtain:

$$\| \lambda_{k+1} - \lambda_* \|_2^2 \leq \| \lambda_k - \lambda_* \|_2^2 - 2\alpha_k g_k^T(\lambda_k - \lambda_*) + \alpha_k^2 \| g_k \|_2^2 . \qquad (3.1b)$$

Inequality 3.1c:

If λ_* is an optimal solution, the condition (2) is satisfied:

$$f_* \leq \max_{0 \leq j \leq \min\{k, M-1\}} f(\lambda_{k-j}) + \gamma(\lambda_* - \lambda_k)^T g_k + \eta_k$$

$$\max_{0 \leq j \leq \min\{k, M-1\}} f(\lambda_{k-j}) - f_* + \eta_k \ \geq \ \gamma(\lambda_k - \lambda_*)^T g_k$$

$$(\lambda_k - \lambda_*)^T g_k \leq \frac{1}{\gamma}\left(\max_{0 \leq j \leq \min\{k, M-1\}} f(\lambda_{k-j}) - f_* + \eta_k \right) \qquad (3.1c)$$

Theorem: *Consider the convex minimization problem D, and suppose an optimal solution λ_* exists. Suppose further that we apply the spectral projected subgradient $SPS2$ with the additional assumption that there exists $G > 0$ such that $\| g \|^2 \leq G$ for all $g \in \partial f(\lambda)$ and any λ in the set $\{\lambda / \| \lambda - \lambda_* \leq \| \lambda_0 - \lambda_* \|\}$. Then*

$$\lim_{k \to \infty} f_k^{best} - f_* < \epsilon.$$

Proof. Combining inequalities (3.1b) and (3.1c),

$$\| \lambda_{k+1} - \lambda_* \|_2^2 \leq \| \lambda_k - \lambda_* \|_2^2 - \frac{2\alpha_k}{\gamma}\left(\max_{0 \leq j \leq \min\{k, M-1\}} f(\lambda_{k-j}) - f_* + \eta_k \right) + \alpha_k^2 \| g_k \|_2^2) \qquad (3.1d)$$

Applying (3.1d) recursively, we have:

$$\begin{aligned}\| \lambda_{k+1} - \lambda_* \|_2^2 &\leq \| \lambda_1 - \lambda_* \|_2^2 - \frac{2}{\gamma} \sum_{i=1}^{k} \alpha_i (\max_{0 \leq j \leq \min\{i, M-1\}, i=1,\ldots,k} f(\lambda_{i-j}) - f_* + \eta_i) \\ &+ \sum_{i=1}^{k} \alpha_i^2 \| g_i \|_2^2 \\ &= \| \lambda_1 - \lambda_* \|_2^2 - \frac{2}{\gamma} \sum_{i=1}^{k} \alpha_i (\max_{0 \leq j \leq \min\{i, M-1\}, i=1,\ldots,k} f(\lambda_{i-j}) - f_*) \\ &- \frac{2}{\gamma} \sum_{i=1}^{k} \alpha_i \eta_i + \sum_{i=1}^{k} \alpha_i^2 \| g_i \|_2^2\end{aligned}$$

Using $\| \lambda_{k+1} - \lambda_* \|_2^2 \geq 0$ we have:

$$\frac{2}{\gamma} \sum_{i=1}^{k} \alpha_i (\max_{0 \leq j \leq \min\{i, M-1\}, i=1,\ldots,k} f(\lambda_{i-j}) - f_*) \leq \| \lambda_1 - \lambda_* \|_2^2 - \frac{2}{\gamma} \sum_{i=1}^{k} \alpha_i \eta_i + \sum_{i=1}^{k} \alpha_i^2 \| g_i \|_2^2$$

$$\max_{0 \leq j \leq \min\{i, M-1\}, i=1,\ldots,k} f(\lambda_{i-j}) - f_* \leq (\frac{\gamma}{2}) \frac{\| \lambda_1 - \lambda_* \|_2^2 - \frac{2}{\gamma} \sum_{i=1}^{k} \alpha_i \eta_i + \sum_{i=1}^{k} \alpha_i^2 \| g_i \|_2^2}{\sum_{i=1}^{k} \alpha_i} \tag{3.1e}$$

Combining (3.1a) with (3.1e) we have the inequality:

$$f_{best}^k - f(\lambda_*) \leq (\frac{\gamma}{2}) \frac{\| \lambda_1 - \lambda_* \|_2^2 - \frac{2}{\gamma} \sum_{i=1}^{k} \alpha_i \eta_i + \sum_{i=1}^{k} \alpha_i^2 \| g_i \|_2^2}{\sum_{i=1}^{k} \alpha_i}$$

Finally, because $\| \lambda_1 - \lambda_* \|_2^2$ is constant called it R and using the assumption $\| g_k \|_2 \leq G$, we obtain the basic inequality:

$$f_{best}^k - f_* \leq (\frac{\gamma}{2}) \frac{R - \frac{2}{\gamma} \sum_{i=1}^{k} \alpha_i \eta_i + G^2 \sum_{i=1}^{k} \alpha_i^2}{\sum_{i=1}^{k} \alpha_i} \tag{3.1f}$$

The inequality (3.1f) can be written as:

$$f_{best}^k - f_* \leq (\frac{\gamma}{2}) [\frac{R^2 - \frac{2}{\gamma} \sum_{i=1}^{k} \alpha_i \eta_i}{\sum_{i=1}^{k} \alpha_i}] + [(\frac{\gamma}{2}) \frac{G^2 \sum_{i=1}^{k} \alpha_i^2}{\sum_{i=1}^{k} \alpha_i}] \qquad (3.1f)$$

In the equation (3.1f) the sequence $(\frac{\gamma}{2})[\frac{R^2-\frac{2}{\gamma}\sum_{i=1}^k \alpha_i\eta_i}{\sum_{i=1}^k \alpha_i}]$ converges to zero as $k \to \infty$, since the numerator converges and the denominator grows without bound. In other hand if the sequence α_k converges to zero and is non-summable, then $[(\frac{\gamma}{2})\frac{G^2\sum_{i=1}^k \alpha_i^2}{\sum_{i=1}^k \alpha_i}]$ converges to zero, which implies the subgradient method converges (in the sense $f_k^{best} \to f_*$). To show this, let $\epsilon > \gamma\delta > 0$ and $\delta > 0$, $\gamma > 0$ then, there exists an integer N_1 such that $\alpha_i \leq \epsilon/G^2$ for all $i > N_1$. There also exists an integer N_2 such that:

$$\sum_{i=1}^{N_1} \alpha_i \geq \frac{1}{\delta}(G^2 \sum_{i=1}^{N_2} \alpha_i^2)$$

since $\sum_{i=1}^{\infty} \alpha_i = \infty$. Let $N = \max N_1, N_2$. Then for $k > N$, we have:

$$\begin{aligned}
(\frac{\gamma}{2})\frac{G^2\sum_{i=1}^k \alpha_i^2}{\sum_{i=1}^k \alpha_i} &\leq (\frac{\gamma}{2})\frac{G^2\sum_{i=1}^{N_1} \alpha_i^2}{\sum_{i=1}^k \alpha_i} + (\frac{\gamma}{2})\frac{G^2\sum_{i=N_1+1}^k \alpha_i^2}{\sum_{i=1}^{N_1} \alpha_i + \sum_{i=N_1+1}^k \alpha_i} \\
&\leq (\frac{\gamma}{2})\frac{G^2\sum_{i=1}^{N_1} \alpha_i^2}{(\frac{1}{\delta})(G^2\sum_{i=1}^{N_1} \alpha_i^2)} + (\frac{\gamma}{2})\frac{G^2\sum_{i=N_1+1}^k \alpha_i(\frac{\delta}{G^2})}{\sum_{i=1}^{N_1} \alpha_i + \sum_{i=N_1+1}^k \alpha_i} \\
&\leq (\frac{\gamma}{2})\frac{1}{\frac{1}{\delta}} + (\frac{\gamma}{2})\delta\frac{\sum_{i=N_1+1}^k \alpha_i}{\sum_{i=1}^{N_1} \alpha_i + \sum_{i=N_1+1}^k \alpha_i} \\
&\leq (\frac{\gamma\delta}{2}) + (\frac{\gamma\delta}{2})\frac{\sum_{i=N_1+1}^k \alpha_i}{\sum_{i=N_1+1}^k \alpha_i} \\
&\leq \epsilon. \quad \diamond
\end{aligned}$$

4 Final Remarks

We study the convergence properties of the spectral projected subgradient method and proved convergence under some mild assumptions. The combination of the SPG1 approach, some conditions over spectral step and the non-monotone line search allowed to establish the convergence of the spectral projected subgradient.

Acknowledgment. The authors thank Marcos Raydan and José Mario Martínez for their constructive suggestions.

References

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

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

[3] D. BERTSIMAS and J. N. TSITSIKLIS. *Introduction to Linear Optimization*. Athena Scientific, Belmont, Massachusetts, 1997.

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

[5] E. G. BIRGIN, J. M. MARTINEZ, and M. RAYDAN. Algorithm 813: SPG-software for convex-constrained optimization. *ACM Transactions on Mathematical Software*, 27:340–349, 2001.

[6] S. BOYD and A. MUTAPCIC. Subgradient Methods, 2007. Notes for EE364b *Stanford University*.

[7] A. CREMA, M. LORETO, and M. RAYDAN. Spectral projected subgradient with a momentum term for the lagrangean dual approach. *Computers and Operation Research*, 34:3174–3186, 2007.

[8] W. LA CRUZ, J. M. MARTINEZ, and M. RAYDAN. Spectral residual method without gradient information for solving large-scale nonlinear systems. *Math. of Comp.*, 75:1449–1466, 2006.

[9] Y. H. DAi and L. Z. LIAO. R-linear convergence of the Barzilai-Borwein gradient method. *IMA J. Numer. Anal.*, 22:1–10, 2002.

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

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

[12] A. M. GEOFFRION. Lagrangean relaxation for integer programing. *Mathematical Programming Study*, 2:82–114, 1974.

[13] A. A. GOLDSTEIN. Convex programming in Hilbert space. *Bull. Amer. Math. Soci.*, 70:709–710, 1964.

[14] L. GRIPPO, F. LAMPARIELLO, and S. LUCIDI. A nonmonotone line search technique for Newton's method. *SIAM J. Numer. Anal.*, 23:707–716, 1986.

[15] M. HELD, P. WOLFE, and H. CROWDER. Validation of subgradient optimization. *Math. Programming*, 6:62–88, 1974.

[16] E. S. LEVITIN and B. T. POLYAK. Constrained minimization problems. *USSR Computl. Math. Mathl. Physics*, 6:1–50, 1966.

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