ISTITUTO DI ANALISI DEI SISTEMI ED INFORMATICA
CONSIGLIO NAZIONALE DELLE RICERCHE

L. Grippo, M. Sciandrone

ON THE CONVERGENCE OF THE BLOCK NONLINEAR GAUSS-SEIDEL METHOD UNDER CONVEX CONSTRAINTS

R. 467 Settembre 1998

Luigi Grippo - Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza", via Buonarroti 12 - 00185 Roma, Italy. Email : grippo@dis.uniroma1.it.

Marco Sciandrone - Istituto di Analisi dei Sistemi ed Informatica del CNR, viale Manzoni 30 - 00185 Roma, Italy. Email : sciandro@iasi.rm.cnr.it.

This work was partially supported by Agenzia Spaziale Italiana, Roma, Italy.

Istituto di Analisi dei Sistemi ed Informatica, CNR
viale Manzoni 30
00185 ROMA, Italy

tel. ++39-06-77161
fax ++39-06-7716461
email: iasi@iasi.rm.cnr.it
URL: http://www.iasi.rm.cnr.it

Abstract

In this paper we state some new convergence results on the minimization version of the block-nonlinear Gauss-Seidel (GS) method for problems with feasible set defined as the Cartesian product of closed convex sets. First we derive some general properties of the limit points produced by the GS method. Using these results we show that in the case of a two-block decomposition, the method is globally convergent towards critical points of the constrained problem even in the absence of any convexity assumption on the objective function f. Then we analyze the convergence properties of the method with a $m-$block decomposition, under convexity assumptions on f. We show that the GS method is globally convergent both when f is componentwise strictly pseudo-convex with respect to $m-2$ blocks, and when f is pseudoconvex and has bounded level sets in the feasible region. We also discuss a constrained version of a counterexample of Powell, which shows that the GS method may fail for $m \geq 3$ when these assumptions are not satisfied. Finally, in the general case of nonconvex objective function and arbitrary decomposition, we prove the global convergence of a proximal point modification of the GS method.

1. Introduction

Consider the problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in X \subseteq R^n, \end{aligned} \tag{1}$$

where $f : R^n \to R$ is a continuously differentiable function and the feasible set X is the Cartesian product of closed, nonempty and convex subsets $X_i \subseteq R^{n_i}$, for $i = 1, \ldots, m$, that is:

$$X = X_1 \times X_2 \times \ldots \times X_m,$$

with $n_1 + n_2 + \ldots n_m = n$.

If the vector $x \in R^n$ is partitioned into m component vectors $x_i \in R^{n_i}$, then the minimization version of the *block-nonlinear Gauss-Seidel* (GS) *method* for the solution of (1) is defined by the iteration:

$$x_i^{k+1} = \arg \min_{y_i \in X_i} f(x_1^{k+1}, \ldots, x_{i-1}^{k+1}, y_i, x_{i+1}^k, \ldots, x_m^k), \tag{2}$$

which updates in sequence the components of x, starting from a given initial point $x^0 \in R^n$.

It is known that, in general, the GS method may not converge, in the sense that it may produce sequences with limit points that are not critical points of the problem. Some well-known examples of this behaviour have been described by Powell [8], with reference to the coordinate method for unconstrained problems, that is to the case $m = n$ and $X = R^n$.

Convergence results for the block GS method have been given under suitable convexity assumptions, both in the unconstrained and in the constrained case, in a number of works (see e.g., [5], [1], [6], [11], [3], [7], [9], [4]).

In the present paper, by extending some of the results established in the unconstrained case, we prove new convergence results of the GS method when applied to constrained problems.

More specifically, first we derive some general properties of the limit points of the partial updates generated by the GS method and we show that each of these points is a critical point at least with respect to two consecutive components in the given ordering. This is shown by proving that the global minimum value in a component subspace is lower than the function value obtainable through a convergent Armijo-type linesearch along a suitably defined feasible direction.

As a consequence of these results, we get a simple proof of the fact that in case of a two block decomposition the GS method (2Block GS method) converges to critical points of problem (1), even in the absence of any convexity assumption on f. The same proof technique is also used for establishing the convergence of a modified version of the 2Block GS method, where the global minimization is performed only with respect to one block of variables.

Then we consider the convergence properties of the GS method for the general case of a $m-$block decomposition under convexity assumptions on the objective function. We show that the GS method is globally convergent both when

(i) f is componentwise strictly pseudoconvex with respect to $m - 2$ blocks

and when

(ii) f is pseudoconvex and has bounded level sets in the feasible region.

In case (i) we get a generalization of well known convergence results ([6], [3]); in case (ii) we extend to the constrained case the convergence results given in [10] for the cyclic coordinate method and in [4] for the unconstrained block GS method.

Using a constrained version of a Powell's counterexample we show that non convergence may arise for nonconvex function, when $m \geq 3$ and the preceding assumptions are not satisfied.

Finally, in the general case of arbitrary decomposition, we consider the proximal point modification of the GS method defined by

$$x_i^{k+1} = \arg\min_{y_i \in X_i} \{f(x_1^{k+1}, \dots, y_i, \dots, x_m^k) + \frac{1}{2}\tau_i \|y_i - x_i^k\|^2\},$$

where $\tau_i > 0$, and we establish the global convergence of this method without any convexity assumption on the objective function.

Notation. We suppose that the vector $x \in R^n$ is partitioned into component vectors $x_i \in R^{n_i}$, as $x = (x_1, x_2, \dots, x_m)$. In correspondence to this partition, the function value $f(x)$ is also indicated by $f(x_1, x_2, \dots, x_m)$ and, for $i = 1, 2, \dots, m$ the partial gradient of f with respect to x_i, evaluated at x, is indicated by $\nabla_i f(x) = \nabla_i f(x_1, x_2, \dots, x_m) \in R^{n_i}$.

A *critical point* for Problem (1) is a point $\bar{x} \in X$ such that

$$\nabla f(\bar{x})^T (y - \bar{x}) \geq 0 \quad \text{for every} \quad y \in X, \tag{3}$$

where $\nabla f(x) \in R^n$ denotes the gradient of f at x. If both $\bar{x}$ and y are are partitioned into component vectors, it easily seen that $\bar{x} \in X$ is a critical point for Problem (1) if and only if for all $i = 1, \dots, m$ we have:

$$\nabla_i f(\bar{x})^T (y_i - \bar{x}_i) \geq 0 \quad \text{for every} \quad y_i \in X_i. \tag{4}$$

We denote by $\mathcal{L}_X$ the level set of f relative to X, corresponding to a given point $x^0 \in X$, that is:

$$\mathcal{L}_X := \{x \in X : f(x) \leq f(x^0)\}.$$

Finally we indicate by $\|\cdot\|$ the Euclidean norm (on the appropriate space).

2. A line search algorithm

In this section we recall some well-known properties of an Armijo-type linesearch algorithm along a feasible direction, which will be used in the sequel in our convergence proofs.

Let $\{z^k\}$ be a given sequence in X, and suppose that z^k is partitioned as

$$z^k = (z_1^k, \dots, z_m^k)$$

with $z_i^k \in X_i$ for $i = 1, \dots, m$. Let us choose an index $i \in \{1, \dots, m\}$ and assume that for all k we can compute search directions

$$d_i^k = w_i^k - z_i^k \qquad \text{with } w_i^k \in X_i, \tag{5}$$

such that the following assumption holds.

Assumption 1. *Let $\{d_i^k\}$ be the sequence of search directions defined by* (5)*. Then:*

(i) *there exists a number $M > 0$ such that $\|d_i^k\| \leq M$ for all k;*

(ii) *we have $\nabla_i f(z^k)^T d_i^k < 0$ for all k.*

An Armijo-type linesearch algorithm can be described as follows.

Line search algorithm (LS)

Data: $\gamma_i \in (0,1)$, $\delta_i \in (0,1)$.

Compute $\alpha_i^k = \max_j\{\delta_i^j : j = 0, 1, \ldots\}$ such that

$$f(z_1^k, \ldots, z_i^k + \alpha_i^k d_i^k, \ldots, z_m^k) \leq f(z^k) + \gamma_i \alpha_i^k \nabla_i f(z^k)^T d_i^k. \quad \square \tag{6}$$

In the next proposition we state some well-known properties of Algorithm LS. It is important to observe that in what follows we assume that $\{z^k\}$ is a given sequence that may not depend on Algorithm LS, in the sense that z_i^{k+1} may be not the result of a linesearch along d_i^k. However this has no substantial effect in the convergence proof, which can easily be deduced from [1].

Proposition 2.1. *Let $\{z^k\}$ be a sequence of points in X and let $\{d_i^k\}$ be a sequence of directions such that Assumption* 1 *is satisfied. Let α_i^k be computed by means of Algorithm LS. Then:*

(i) *there exists a finite integer j such that $\alpha_i^k = \delta_i^j$ satisfies the acceptability condition* (6)*;*

(ii) *if $\{z^k\}$ converges to $\bar{z}$ and:*

$$\lim_{k\to\infty} f(z^k) - f(z_1^k, \ldots, z_i^k + \alpha_i^k d_i^k, \ldots, z_m^k) = 0, \tag{7}$$

then we have:

$$\lim_{k\to\infty} \nabla_i f(z^k)^T d_i^k = 0. \tag{8}$$

Proof. In order to prove the first assertion, let us assume, by contradiction, that $\nabla_i f(z^k)^T d_i^k < 0$ and that condition (6) is violated for every $j \geq 0$, so that

$$\frac{f(z_1^k, \ldots, z_i^k + \delta_i^j d_i^k, \ldots, z_m^k) - f(z^k)}{\delta_i^j} > \gamma_i \nabla_i f(z^k)^T d_i^k.$$

Then, taking limits for $j \to \infty$ we obtain:

$$\nabla_i f(z^k)^T d_i^k \geq \gamma_i \nabla_i f(z^k)^T d_i^k,$$

which contradicts the assumptions $\nabla_i f(z^k)^T d_i^k < 0$ and $\gamma_i < 1$.

Now, in order to prove the second assertion, suppose that z^k converges to $\bar{z}$ and that condition (7) holds.

Reasoning by contradiction, as the sequence $\nabla_i f(z^k)^T d_i^k$ is bounded (because of the fact that both the sequences $\{z^k\}$ and $\{d_i^k\}$ are bounded), we can assume that there exists a subsequence $\{z^k\}_K$ such that:

$$\lim_{k\to\infty, k\in K} \nabla_i f(z^k)^T d_i^k = -\mu < 0, \tag{9}$$

for some $\mu > 0$. Since $\{d_i^k\}$ is bounded we can extract a further subsequence $\{d_i^k\}_{K_1}$ with $K_1 \subseteq K$ such that

$$\lim_{k\to\infty, k\in K_1} d_i^k = \bar{d}_i, \tag{10}$$

for some $\bar{d}_i \in R^{n_1}$. Thus, by (9), (10) and the continuity of $\nabla_i f$ we can write

$$\lim_{k\to\infty, k\in K_1} \nabla_i f(z^k)^T d_i^k = \nabla_i f(\bar{z})^T \bar{d}_i = -\mu. \tag{11}$$

Now, as α_i^k satisfies condition (6) and $\nabla_i f(z^k)^T d_i^k < 0$, we have:

$$f(z^k) - f(z_1^k, \dots, z_i^k + \alpha_i^k d_i^k, \dots, z_m^k) \ge -\gamma_i \alpha_i^k \nabla_i f(z^k)^T d_i^k.$$

Then, taking limits for $k \to \infty$ and recalling assumption (7), we obtain

$$\lim_{k\to\infty} \alpha_i^k \nabla_i f(z^k)^T d_i^k = 0, \tag{12}$$

so that, by (11), we have:

$$\lim_{k\to\infty, k\in K_1} \alpha_i^k = 0, \tag{13}$$

and therefore, for sufficiently large $k \in K_1$, we have: $\alpha_i^k < 1$, so that α_i^k/δ_i violates the condition at Step 1 of Algorithm LS. Hence, using the Mean Value Theorem we can write:

$$f(z^k) + \alpha_i^k \nabla_i f(\tilde{z}^k)^T d_i^k > f(z^k) + \gamma_i \alpha_i^k \nabla_i f(z^k)^T d_i^k \tag{14}$$

where

$$\tilde{z}^k := (z_1^k, \dots, z_i^k + \lambda_i^k \frac{\alpha_i^k}{\delta_i} d_i^k, \dots, z_m^k), \tag{15}$$

and $\lambda_i^k \in (0,1)$. As $\{z^k\}$ converges to $\bar{z}$, by (13) and the boundedness of $\{d_i^k\}$ we have that

$$\lim_{k\to\infty, k\in K_1} \tilde{z}^k = \bar{z}.$$

Now by (14) we can write:

$$\nabla_i f(\tilde{z}^k)^T d_i^k > \gamma_i \nabla_i f(z^k)^T d_i^k. \tag{16}$$

Then, taking limits for $k \to \infty$ and $k \in K_1$, by (11) we obtain: $-\mu \ge -\gamma_i \mu$, which contradicts the assumption $\gamma_i \in (0,1)$. ∎

3. Preliminary results

In this section we derive some properties of the GS method that are at the basis of some of our convergence results. First we state the $m-$block GS method in the following form.

GS Method

Data. $x^0 \in X$.

Step 0. Set $k = 0$.

Step 1. For $i = 1, \dots, m$:

compute

$$x_i^{k+1} = \arg\min_{y_i \in X_i} f(x_1^{k+1}, \ldots, y_i, \ldots, x_m^k). \qquad (17)$$

Step 2. Set

$$x^{k+1} = \left(x_1^{k+1}, \ldots, x_m^{k+1}\right),$$

$k = k + 1$ and go to Step 1. $\square$

We assume in the sequel that the updating rule (17) is well defined; hovewer we do not assume that the minimum is uniquely attained. We consider, for all k, the partial updates introduced by the GS method by defining the following vectors belonging to X:

$$\begin{array}{ll} w(k,0) = x^k & \\ w(k,i) = (x_1^{k+1}, \ldots, x_{i-1}^{k+1}, x_i^{k+1}, x_{i+1}^k, \ldots, x_m^k) & \quad i = 1, \ldots, m-1 \\ w(k,m) = x^{k+1}. & \end{array}$$

For convenience we set also

$$w(k, m+1) = w(k+1, 1).$$

By construction, for each $i \in \{1, \ldots, m\}$, it follows from (17) that $w(k,i)$ is the constrained global minimizer of f in the $i-$th component subspace, and therefore it satisfies the necessary optimality condition:

$$\nabla_i f(w(k,i))^T (y_i - x_i^{k+1}) \geq 0 \quad \text{for every } y_i \in X_i. \qquad (18)$$

We can state the following propositions.

Proposition 3.1. *Suppose that in correspondence to a given $i \in \{0, \ldots, m\}$ the sequence $\{w(k,i)\}$ admits a limit point $\bar{w}$. Then, for every $j \in \{0, \ldots, m\}$ we have*

$$\lim_{k \to \infty} f(w(k,j)) = f(\bar{w}).$$

Proof. Let us consider an infinite subset $K \subseteq \{0, 1, \ldots, \}$ and an index $i \in \{0, \ldots, m\}$ such that the subsequence $\{w(k,i)\}_K$ converges to a point $\bar{w}$. By the instructions of the algorithm we have

$$f(w(k+1,i)) \leq f(w(k,i)). \qquad (19)$$

Then, the continuity of f and the convergence of $\{w(k,i)\}_K$ imply that the sequence $\{f(w(k,i))\}$ has a subsequence converging to $f(\bar{w})$. As $\{f(w(k,i))\}$ is non increasing, this, in turn, implies that $\{f(w(k,i))\}$ is bounded from below and converges to $f(\bar{w})$. Then, the assertion follows immediately from the fact that

$$f(w(k+1,i)) \leq f(w(k,j)) \leq f(w(k,i)) \qquad \text{for } 0 \leq j \leq i,$$

and

$$f(w(k+2,i)) \leq f(w(k+1,j)) \leq f(w(k+1,i)) \qquad \text{for } i < j \leq m.$$

∎

Proposition 3.2. *Suppose that in correspondence to a given $i \in \{1, \dots, m\}$ the sequence $\{w(k,i)\}$ admits a limit point $\bar{w}$. Then we have*

$$\nabla_i f(\bar{w})^T (y_i - \bar{w}_i) \geq 0 \qquad \text{for every } y_i \in X_i \tag{20}$$

and moreover

$$\nabla_{i^*} f(\bar{w})^T (y_{i^*} - \bar{w}_{i^*}) \geq 0 \qquad \text{for every } y_{i^*} \in X_{i^*}, \tag{21}$$

where $i^* = i(\text{mod} m) + 1$.

Proof. Let $\{w(k,i)\}_K$ be a subsequence converging to $\bar{w}$. ¿From (18), taking into account the continuity assumption on $\nabla_i f$, we get immediately (20).
In order to prove (21), suppose first $i \in \{1, \dots, m-1\}$, so that $i^* = i+1$. Reasoning by contradiction, let us assume that there exists a vector $\tilde{y}_{i+1} \in X_{i+1}$ such that

$$\nabla_{i+1} f(\bar{w})^T (\tilde{y}_{i+1} - \bar{w}_{i+1}) < 0. \tag{22}$$

Then, letting

$$d^k_{i+1} = \tilde{y}_{i+1} - w(k,i)_{i+1} = \tilde{y}_{i+1} - x^k_{i+1},$$

as $\{w(k,i)\}_K$ is convergent, we have that the sequence $\{d^k_{i+1}\}_K$ is bounded. Recalling (22) and taking into account the continuity assumption on $\nabla_i f$ it follows that there exists a subset $K_1 \subseteq K$ such that

$$\nabla_{i+1} f(w(k,i))^T d^k_{i+1} < 0 \qquad \text{for all } k \in K_1,$$

therefore the sequences $\{w(k,i)\}_{K_1}$ and $\{d^k_{i+1}\}_{K_1}$ are such that Assumption 1 holds, provided that we identify $\{z^k\}$ with $\{w(k,i)\}_{K_1}$.
Now, for all $k \in K_1$ suppose that we compute α^k_{i+1} by means of Algoritm LS; then we have

$$f(x^{k+1}_1, \dots, x^{k+1}_i, x^k_{i+1} + \alpha^k_{i+1} d^k_{i+1}, \dots, x^k_m) \leq f(w(k,i)).$$

Moreover, as $x^k_{i+1} \in X_{i+1}$, $x^k_{i+1} + d^k_{i+1} \in X_{i+1}$, $\alpha^k_{i+1} \in (0,1]$, and X_{i+1} is convex, it follows that

$$x^k_{i+1} + \alpha^k_{i+1} d^k_{i+1} \in X_{i+1}.$$

Therefore, recalling that

$$f(w(k,i+1)) = \min_{y_{i+1} \in X_{i+1}} f(x^{k+1}_1, \dots, x^{k+1}_i, y_{i+1}, \dots, x^k_m),$$

we can write:

$$f(w(k,i+1)) \leq f(x^{k+1}_1, \dots, x^{k+1}_i, x^k_{i+1} + \alpha^k_{i+1} d^k_{i+1}, \dots, x^k_m) \leq f(w(k,i)). \tag{23}$$

By Proposition 3.1 we have that the sequences $\{f(w(k,j))\}$ are convergent to a unique limit for all $j \in \{0 \dots, m\}$, and hence we obtain

$$\lim_{k \to \infty, k \in K_1} f(w(k,i)) - f(x^{k+1}_1, \dots, x^{k+1}_i, x^k_{i+1} + \alpha^k_{i+1} d^k_{i+1}, \dots, x^k_m) = 0.$$

Then, invoking Proposition 2.1, where we identify $\{z^k\}$ with $\{w(k,i)\}_{K_1}$, it follows that

$$\nabla_{i+1} f(\bar{w})^T (\tilde{y}_{i+1} - \bar{w}_{i+1}) = 0,$$

which contradicts (22), so that we have proved that (21) holds when $i \in \{1, \dots, m-1\}$.

When $i = m$, so that $i^* = 1$, we can repeat the same reasonings noting that $w(k, m+1) = w(k+1, 1)$. ∎

The preceding result implies, in particular, that every limit point of the sequence $\{x^k\}$ generated by the GS method is a critical point with respect to the components x_1 and x_m in the prefixed ordering. This is formally stated in the next proposition.

Proposition 3.3. *Let $\{x^k\}$ be the sequence generated by the* GS *method and suppose that there exists a limit point $\bar{x}$. Then we have:*

$$\nabla_1 f(\bar{x})^T (y_1 - \bar{x}_1) \geq 0 \qquad \text{for every } y_1 \in X_1 \tag{24}$$

and

$$\nabla_m f(\bar{x})^T (y_m - \bar{x}_m) \geq 0. \qquad \text{for every } y_m \in X_m. \tag{25}$$

∎

4. The two-block GS method

Let us consider the problem:

$$\begin{array}{ll} \text{minimize} & f(x) = f(x_1, x_2), \\ x \in X_1 \times X_2 & \end{array} \tag{26}$$

under the assumptions stated in Section 1. We note that in many cases a two-block decomposition can be useful since it may allow us to employ parallel techniques for solving one subproblem. As an example, given a function of the form

$$f(x) = \psi_1(x_1) + \sum_{i=2}^{N} \psi_1(x_1)\phi_i(x_i),$$

if we decompose the problem variables into the two blocks x_1 and $(x_2, \dots, x_N)$, then once x_1 is fixed, the objective function can be minimized in parallel with respect to the components x_i for $i = 2, \dots, N$.

When $m = 2$, a convergence proof for the GS method (2Block GS method) in the unconstrained case was given in [4]. Here the extension to the constrained case is an immediate consequence of Proposition 3.3.

Proposition 4.1. *Let $\{x^k\}$ be the sequence generated by the* 2Block GS *method. Then:*

(i) *every limit point $\bar{x}$ of $\{x^k\}$ is a critical point of Problem* (1)*;*

(ii) *if $\mathcal{L}_X$ is compact the sequence $\{x^k\}$ has at least one limit point, which is a critical point of Problem* (26)*;*

(iii) *if f is pseudoconvex on X every limit point of $\{x^k\}$ is a global minimizer of Problem* (26). ∎

We can note that the 2Block GS method involves, at each iteration, two global minimization subproblems. However, on the basis of the proof of Proposition 3.2, we can define a modified version of the method that requires, at each iteration, the solution of one global minimization subproblem in a subspace and the computation of a critical point in the other subspace. In fact, in the proof of Proposition 3.2, we use only the fact that $w(k,i)$ is a critical point in the $i-$th component subspace such that $f(w(k,i)) \le f(w(k,i-1))$. Taking this into account, we can define the following modified algorithm, whose convergence properties are exactly the same stated in the preceding proposition for the 2Block GS method.

Modified 2Block GS method

Step 0. Given $x^0 \in R^n$, set $k = 0$.

Step 1a. Set

$$x_1^{k+1} = \arg\min_{y_1 \in X_1} f(y_1, x_2^k); \tag{27}$$

Step 1b. *Determine* x_2^{k+1} such that $f(x_1^{k+1}, x_2^{k+1}) \le f(x_1^{k+1}, x_2^k)$ and

$$\nabla_2 f(x_1^{k+1}, x_2^{k+1})^T (y_2 - x_2^{k+1}) \ge 0 \qquad \text{for all } y_2 \in X_2. \tag{28}$$

Step 2. Set

$$x^{k+1} = \left(x_1^{k+1}, x_2^{k+1}\right),$$

$k = k + 1$ and go to Step 1. $\square$

We note that the modified algorithm can be useful when a two-block decomposition yields one convex subproblem, while the other subproblem is non-convex and has to be solved with an ordinary local optimization technique.

5. Convergence of the block GS method under convexity assumptions

In this section we analyse the convergence properties of the *block nonlinear Gauss-Seidel* method in the case of arbitrary decomposition. In particular we show that in this case the global convergence of the method can be ensured

- assuming the strict componentwise convexity of the objective function with respect to $m-2$ components,
- or assuming that the objective function is pseudoconvex and has bounded level sets.

First we state the following proposition, whose proof requires only minor adaptations of the arguments used, for instance, in [6] and [3].

Proposition 5.1. *Suppose that f is a strictly pseudoconvex function of x_i when the other components are held constant. Let $\{y^k\}$ be a sequence of points in X converging to some $\bar{y} \in X$ and let $\{v^k\}$ be a sequence of vectors whose components are defined as follows:*

$$v_j^k = \begin{cases} y_j^k & if\ j \ne i \\ \arg\min_{\xi \in X_i} f(y_1^k, \ldots, \xi, \ldots, y_m^k) & if\ j = i \end{cases}$$

Then, if

$$\lim_{k\to\infty} f(y^k) - f(v^k) = 0,$$

we have

$$\lim_{k\to\infty} \|v_i^k - y_i^k\| = 0.$$

Proof. Reasoning by contradiction, let us suppose that there exist an infinite subset K such that $\{y^k\}_K$ converges to $\bar{y}$ and we have

$$\|v_i^k - y_i^k\| > \beta > 0 \quad \text{for all} \quad k \in K. \tag{29}$$

For $k \in K$, let $s^k = \dfrac{v^k - y^k}{\|v^k - y^k\|}$ and consider the point

$$\tilde{v}^k := y^k + \beta s^k = (1 - \frac{\beta}{\|v^k - y^k\|})y^k + \frac{\beta}{\|v^k - y^k\|}v^k.$$

The point $\tilde{v}^k$ belongs to X, since y^k,$v^k \in X$, $\frac{\beta}{\|v^k-y^k\|} < 1$ and X is convex.
As $\|s^k\| = 1$, we can find an infinite subset $K_1 \subseteq K$ such that:

$$\lim_{k\to\infty, k\in K_1} y^k = \lim_{k\to\infty, k\in K_1} (y_1^k, \dots, y_m^k) = (\bar{y}_1, \dots, \bar{y}_i, \dots, \bar{y}_m) = \bar{y}$$

and:

$$\lim_{k\to\infty, k\in K_1} \tilde{v}^k = \lim_{k\to\infty, k\in K_1} (y_1^k, \dots, y_i^k + \beta s^k, \dots, y_m^k) = (\bar{y}_1, \dots, y_i^*, \dots, \bar{y}_m) = y^*$$

with

$$\|\bar{y}_i - y_i^*\| = \beta > 0.$$

On the other hand, taking into account the strict pseudoconvexity assumption on f with respect to the $i-$th component, we have for all $t \in [0, 1]$

$$f(y^k) \geq f(y_1^k, \dots, (1-t)y_i^k + t(y_i^k + \beta s_i^k), \dots, y_m^k) \geq f(v^k). \tag{30}$$

Suppose now, as stated in the assertion, that

$$\lim_{k\to\infty} f(y^k) - f(v^k) = 0. \tag{31}$$

Then, taking limits in (30), for $k \to \infty$, $k \in K_1$, by (31) and the continuity of f we get

$$f(\bar{y}) = f(\bar{y}_1, \dots, (1-t)\bar{y}_i + ty_i^*, \dots, \bar{y}_m) \quad \text{for all } t \in [0, 1],$$

which contradicts the strict pseudoconvexity assumption. ∎

Then, we can state the following proposition.

Proposition 5.2. *Suppose that the function f is strictly pseudoconvex with respect to x_i for $i = 1, \dots, m-2$. Then, the* GS *method generates an infinite sequence $\{x^k\}$ such that:*

(i) *every limit point $\bar{x}$ of $\{x^k\}$ is a critical point of Problem* (1)*;*

(ii) *if $\mathcal{L}_X$ is compact the sequence $\{x^k\}$ has at least one limit point, which is a critical point of Problem* (1)*;*

(iii) *if f is pseudoconvex on X every limit point of $\{x^k\}$ is a global minimizer of Problem* (1).

Proof. Let us assume that there exists a subsequence $\{x^k\}_K$ such that

$$\lim_{k\to\infty, k\in K} x^k = \bar{x}. \tag{32}$$

From Proposition 3.3 we get

$$\nabla_m f(\bar{x})^T (y_m - \bar{x}_m) \geq 0 \qquad \text{for every } y_m \in X_m. \tag{33}$$

Recalling Proposition 3.1 we can write

$$\lim_{k\to\infty} f(w(k,i)) - f(x^k) = 0 \qquad \text{for } i = 1, \dots, m.$$

Then, by using the strict pseudoconvexity assumption on f and invoking Proposition 5.1, where we identify $\{y^k\}$ with $\{x^k\}_K$ and $\{v^k\}$ with $\{w(k,1)\}_K$, we obtain

$$\lim_{k\to\infty, k\in K} w(k,1) = \bar{x}.$$

By repeated application of Proposition 5.1 to the sequences $\{w(k,i)\}_K$ and $\{w(k,i+1)\}_K$, for $i = 1, \dots, m-2$, we obtain

$$\lim_{k\to\infty, k\in K} w(k,i) = \bar{x} \qquad \text{for } i = 1, \dots, m-2.$$

Proposition 3.2 implies

$$\nabla_i f(\bar{x})^T (y_i - \bar{x}_i) \geq 0 \qquad \text{for every } y_i \in X_i, \quad i = 1, \dots, m-1. \tag{34}$$

Hence, assertion (i) follows from (33) and (34). Assertions (ii) and (iii) are immediate consequences of (i). ∎

In the next proposition we consider the case of pseudoconvex objective function.

Proposition 5.3. *Suppose that f is pseudoconvex and that $\mathcal{L}_X$ is compact. Then, the* GS *method generates an infinite sequence $\{x^k\}$ such that every limit point $\bar{x}$ of $\{x^k\}$ is a global minimizer of f.*

Proof. Consider the partial updates $w(k,i)$, with $i = 0, \dots, m$, defined in Section 3. By definition of $w(k,i)$ we have $f(x^{k+1}) \leq f(w(k,i)) \leq f(x^k)$ for $i = 0, \dots, m$. Then, the points of the sequences $\{w(k,i)\}$, with $i = 0, \dots, m$, belong to the compact set $\mathcal{L}_X$. Therefore, if $\bar{x} \in X$ is a limit point of $\{x^k\}$ we can construct a subsequence $\{x^k\}_K$ such that

$$\lim_{k\to\infty, k\in K} x^k = \bar{x} = \bar{w}^0 \tag{35}$$

$$\lim_{k\to\infty, k\in K} w(k,i) = \bar{w}^i \qquad i = 1, \dots, m, \tag{36}$$

where $\bar{w}^i \in X$ for $i = 1, \dots, m$. We can write

$$w(k,i) = w(k,i-1) + d(k,i) \qquad \text{for } i = 1, \dots, m, \tag{37}$$

where the block components $d_h(k,i) \in R^{n_h}$ of the vector $d(k,i)$, with $h \in \{1, \dots, m\}$, are such that $d_h(k,i) = 0$ if $h \neq i$. Therefore, for $i = 1, \dots, m$, from (35)-(37) we get

$$\bar{w}^i = \bar{w}^{i-1} + \bar{d}^i, \tag{38}$$

where

$$\bar{d}^i = \lim_{k\to\infty, k\in K} d(k,i), \tag{39}$$

and

$$\bar{d}^i_h = 0 \qquad h \neq i. \tag{40}$$

By Proposition 3.1 we have

$$f(\bar{x}) = f(\bar{w}^i) \qquad \text{for } i = 1, \dots, m. \tag{41}$$

From Proposition 3.2 it follows, for $i = 1, \dots, m$,

$$\nabla_i f(\bar{w}^i)^T (y_i - \bar{w}^i_i) \geq 0 \qquad \text{for all } y_i \in X_i \tag{42}$$

and

$$\nabla_{i^*} f(\bar{w}^i)^T (y_{i^*} - \bar{w}^i_{i^*}) \geq 0 \qquad \text{for all } y_{i^*} \in X_{i^*}, \tag{43}$$

where $i^* = i(\mathrm{mod} m) + 1$. Now we prove that, given $j,\ \ell \in \{1, \dots, m\}$ such that

$$\nabla_\ell f(\bar{w}^j)^T (y_\ell - \bar{w}^j_\ell) \geq 0 \qquad \text{for all } y_\ell \in X_\ell, \tag{44}$$

then it follows

$$\nabla_\ell f(\bar{w}^{j-1})^T (y_\ell - \bar{w}^{j-1}_\ell) \geq 0 \qquad \text{for all } y_\ell \in X_\ell. \tag{45}$$

Obviously, (45) holds if $\ell = j$ (see (43)). Therefore, let us assume $\ell \neq j$. By (38)-(40) we have

$$\bar{w}^j = \bar{w}^{j-1} + \bar{d}^j, \tag{46}$$

where $\bar{d}^j_h = 0$ for $h \neq j$. For any given vector $\eta \in R^{n_\ell}$ such that

$$\bar{w}^{j-1}_\ell + \eta \in X_\ell,$$

let us consider the feasible point

$$z(\eta) = \bar{w}^{j-1} + d(\eta),$$

where $d_h(\eta) = 0$ for $h \neq \ell$ and $d_\ell(\eta) = \eta \in R^{n_\ell}$. Then, from (42) and (44), observing that (46) and the fact that $\ell \neq j$ imply

$$\eta = z_\ell(\eta) - \bar{w}^{j-1}_\ell = z_\ell(\eta) - \bar{w}^j_\ell,$$

we obtain

$$\begin{aligned} \nabla f(\bar{w}^j)^T (z(\eta) - \bar{w}^j) &= \nabla f(\bar{w}^j)^T (\bar{w}^{j-1} + d(\eta) - \bar{w}^j) \\ &= \nabla_j f(\bar{w}^j)^T (\bar{w}^{j-1}_j - \bar{w}^j_j) + \nabla_\ell f(\bar{w}^j)^T \eta \\ &= \nabla_j f(\bar{w}^j)^T (\bar{w}^{j-1}_j - \bar{w}^j_j) + \nabla_\ell f(\bar{w}^j)^T (z_\ell(\eta) - \bar{w}^j_\ell) \geq 0. \end{aligned}$$

It follows by the pseudoconvexity of f that

$$f(z(\eta)) \geq f(\bar{w}^j) \qquad \text{for all } \eta \in R^{n_\ell} \text{ such that } \bar{w}_\ell^{j-1} + \eta \in X_\ell.$$

On the other hand, $f(\bar{w}^j) = f(\bar{w}^{j-1})$, and therefore we have:

$$f(z(\eta)) \geq f(\bar{w}^{j-1}) \qquad \text{for all } \eta \in R^{n_\ell} \text{ such that } \bar{w}_\ell^{j-1} + \eta \in X_\ell,$$

which, recalling the definition of $z(\eta)$, implies (45).
Finally, taking into account (43), and using the fact that (44) implies (45), by induction we obtain

$$\nabla_j f(\bar{w}^0)^T (y_j - \bar{w}_j^0) = \nabla_j f(\bar{x})^T (y_j - \bar{x}_j) \geq 0 \qquad \text{for all } y_j \in X_j.$$

Since this is true for every $j \in \{1, \ldots, m\}$, the thesis is proved. ∎

6. A counterexample

In this section we consider a constrained version of a well known counterexample due to Powell [8] which indicates that the results given in the preceding sections are tight in some sense. In fact this example shows that the GS method may cycle indefinitely without converging to a critical point if the number m of blocks is equal to 3 and the objective f is a non convex function, which is componentwise convex but not strictly convex with respect to each component.

The original counterexample of Powell consists in the unconstrained minimization of the function $f : R^3 \to R$, defined by:

$$\begin{aligned} f(x) = \quad & -x_1x_2 - x_2x_3 - x_1x_3 + (x_1 - 1)_+^2 + (-x_1 - 1)_+^2 + \\ & +(x_2 - 1)_+^2 + (-x_2 - 1)_+^2 + (x_3 - 1)_+^2 + (-x_3 - 1)_+^2, \end{aligned} \tag{47}$$

where

$$(t - c)_+^2 = \begin{cases} 0 & \text{if } \ t \leq c \\ (t - c)^2 & \text{if } \ t \geq c \end{cases}$$

Powell showed that, if the starting point x^0 is the point $(-1 - \epsilon, 1 + \frac{1}{2}\epsilon, -1 - \frac{1}{4}\epsilon)$ the steps of the GS method "tend to cycle round six edges of the cube whose vertices are $(\pm 1, \pm 1, \pm 1)$", which are not stationary points of f. It can be easily verified that the level sets of the objective function (47) are not compact; in fact, setting $x_2 = x_3 = x_1$ we have that $f(x) \to -\infty$ as $\|x\| \to \infty$. However, the same behaviour evidenced by Powell is obtained if we consider a constrained problem with the same objective function (47) and a compact feasible set, defined by the box constraints

$$-M \leq x_i \leq M \qquad i = 1, \ldots, 3$$

with $M > 0$ sufficiently large.

In accordance with the results of Section 3, we may note that the limit points of the partial updates generated by the GS method are such that two gradient components are zero.

Non convergence is due to the fact that the limit points associated to consecutive partial updates are distinct because of the fact that the function is not componentwise strictly pseudoconvex; on the other hand, as the function is not pseudoconvex, the limit points of the sequence $\{x^k\}$ are not critical points.

Note that in the particular case of $m = 3$, by Proposition 5.2 we can ensure convergence by requiring only the strict pseudoconvexity of f with respect to one component.

7. A proximal point modification of the GS method

In the preceding sections we have shown that the global convergence of the GS method can be ensured either under suitable convexity assumptions on the objective function or in the special case of a two-block decomposition.

Now, for the general case of nonconvex objective function and arbitrary decomposition, we consider a proximal point modification of the Gauss-Seidel method. Proximal point versions of the GS method have been already considered in the literature (see e.g., [1], [2]), but only under convexity assumptions on f. Here we show that these assumptions are not required if we are interested only in critical points. The algorithm can be described as follows.

PGS method

Data. $x^0 \in X$, $\tau_i > 0$ for $i = 1, \ldots, m$.

Step 0. Set $k = 0$.

Step 1. For $i = 1, \ldots, m$ set:

$$x_i^{k+1} = \arg\min_{y_i \in X_i} \{f(x_1^{k+1}, \ldots, y_i, \ldots, x_m^k) + \frac{1}{2}\tau_i \|y_i - x_i^k\|^2\}. \tag{48}$$

Step 2. Set

$$x^{k+1} = \left(x_1^{k+1}, \ldots, x_m^{k+1}\right),$$

$k = k + 1$ and go to Step 1. $\square$

The convergence properties of the method are given in the following proposition.

Proposition 7.1. *Let $\{x^k\}$ be the sequence generated by the* PGS *method. Then:*

(i) *every limit point $\bar{x}$ of $\{x^k\}$ is a critical point of Problem* (1)*;*

(ii) *if $\mathcal{L}_X$ is compact the sequence $\{x^k\}$ has at least one limit point, which is a critical point of Problem* (1)*;*

(iii) *if f is pseudoconvex on X every limit point of $\{x^k\}$ is a global minimizer of Problem* (1).

Proof. Let us assume that there exists a subsequence $\{x^k\}_K$ such that

$$\lim_{k \to \infty, k \in K} x^k = \bar{x}. \tag{49}$$

Define the vectors

$$\tilde{w}(k, 0) = x^k,$$

$$\tilde{w}(k, i) = (x_1^{k+1}, \ldots, x_i^{k+1}, x_{i+1}^k, \ldots, x_m^k) \qquad \text{for } i = 1, \ldots, m.$$

Then we have:

$$f(\tilde{w}(k, i)) \leq f(\tilde{w}(k, i-1)) - \frac{1}{2}\tau_i \|\tilde{w}(k, i) - \tilde{w}(k, i-1)\|^2, \tag{50}$$

from which it follows

$$f(x^{k+1}) \le f(\tilde{w}(k,i)) \le f(\tilde{w}(k,i-1)) \le f(x^k) \quad \text{for} \quad i = 1,\dots,m. \tag{51}$$

Reasoning as in Proposition 3.1 we obtain

$$\lim_{k\to\infty} f(x^{k+1}) - f(x^k) = 0.$$

Hence, taking limits in (50) for $k \to \infty$ we have

$$\lim_{k\to\infty} \|\tilde{w}(k,i) - \tilde{w}(k,i-1)\| = 0 \qquad i = 1,\dots,m, \tag{52}$$

which implies

$$\lim_{k\to\infty} \tilde{w}(k,i) = \bar{x} \qquad i = 0,\dots,m. \tag{53}$$

Now, for every $j \in \{1,\dots,m\}$, as x_j^{k+1} is generated according to the rule (48), the point

$$\tilde{w}(k,j) = (x_1^{k+1},\dots,x_j^{k+1},\dots,x_m^k)$$

satisfies the optimality condition

$$\left[\nabla_j f(\tilde{w}(k,j)) + \tau_j(\tilde{w}_j(k,j) - \tilde{w}_j(k,j-1))\right]^T (y_j - \tilde{w}_j)(k,j) \ge 0 \quad \forall y_j \in X_j.$$

Then, taking the limit for $k \to \infty$, $k \in K$, recalling (52), (53) and the continuity assumption on ∇f, for every $j \in \{1,\dots,m\}$ we obtain

$$\nabla_j f(\bar{x})^T (y_j - \bar{x}_j) \ge 0 \qquad \text{for all } y_j \in X_j,$$

which proves assertion (i). Assertions (ii) and (iii) follow immediately from (i). ∎

Taking into account the results of Section 5, it follows that if the objective function f is strictly pseudoconvex with respect to some component x_i, with $i \in \{1,\dots,m\}$, then we can set $\tau_i = 0$.

Moreover, reasoning as in the proof of Proposition 5.2, we can obtain the same convergence results if we set $\tau_{m-1} = \tau_m = 0$.

8. Concluding remarks

The results stated in this paper show that the GS method can be employed for large classes of problems which include many interesting applications.

We remark that globally convergent decomposition algorithms for constrained problems can also be obtained by replacing the global minimizations used in the GS method with suitable searches along feasible gradient related directions (see e.g., [3], [4], [7], [9]).

The rationale behind these approaches is again that of ensuring that partial consecutive updates do not converge towards distinct limit points that are not critical points for the problem.

References

[1] D.P. Bertsekas. *Nonlinear programming.* Athena Scientific, Belmond, Massachusetts, 1995.

[2] D.P. Bertsekas and P. Tseng. Partial proximal minimization algorithms for convex programming. *SIAM Journal on Optimization*, 4(3):551–572, 1994.

[3] D.P. Bertsekas and J.N. Tsitsiklis. *Parallel and Distributed Computation.* Prentice-Hall International Editions, Englewood Cliffs, New Jersey, 1989.

[4] L. Grippo and M. Sciandrone. Globally convergent block-coordinate techniques for unconstrained optimization. *Optimization methods and software*, 1998. To appear.

[5] Z.Q. Luo and P. Tseng. On the convergence of the coordinate descent method for convex differentiable minimization. *Journal of Optimization Theory and Applications*, 72:7–35, 1992.

[6] J.M. Ortega and W.C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables.* Academic Press, New York, 1970.

[7] M. Patriksson. Decomposition methods for differentiable optimization problems over Cartesian product sets. *Computational optimization and applications*, 9(1):5–42, 1998.

[8] M.J.D. Powell. On search directions for minimization algorithms. *Mathematical Programming*, 4:193–201, 1973.

[9] P. Tseng. Decomposition algorithm for convex differentiable minimization. *Journal of Optimization Theory and Applications*, 70:109–135, 1991.

[10] N. Zadeh. A note on the cyclic coordinate ascent method. *Management Science*, 16:642–644, 1970.

[11] W. I. Zangwill. *Nonlinear Programming: A Unified Approach.* Prentice Hall, Englewood Cliffs, 1969.