On the convergence of the block nonlinear Gauss-Seidel method under convex constraints.

On the convergence of the block nonlinear Gauss–Seidel method under convex constraints ☆

L. Grippo[a], M. Sciandrone[b], *

[a]*Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza", Via Buonarroti 12-00185 Roma, Italy*
[b]*Istituto di Analisi dei Sistemi ed Informatica del CNR, Viale Manzoni 30-00185 Roma, Italy*

Abstract

We give new convergence results for the block Gauss–Seidel method for problems where the feasible set is the Cartesian product of m closed convex sets, under the assumption that the sequence generated by the method has limit points. We show that the method is globally convergent for $m = 2$ and that for $m > 2$ convergence can be established both when the objective function f is componentwise strictly quasiconvex with respect to $m - 2$ components and when f is pseudoconvex. Finally, we consider a proximal point modification of the method and we state convergence results without any convexity assumption on the objective function. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Nonlinear programming; Algorithms; Decomposition methods; Gauss–Seidel method

1. Introduction

Consider the problem

$$\text{minimize} \quad f(x)$$
$$\text{subject to} \quad x \in X = X_1 \times X_2 \times \cdots \times X_m \subseteq R^n, \quad (1)$$

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 $i = 1, \ldots, m$, with $\sum_{i=1}^{m} n_i = n$. 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:

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

which updates in turn the components of x, starting from a given initial point $x^0 \in X$ and generates a sequence $\{x^k\}$ with $x^k = (x_1^k, \ldots, x_m^k)$.

It is known that, in general, the GS method may not converge, in the sense that it may produce a sequence with limit points that are not critical points of the problem.

Some well-known examples of this behavior have been described by Powell [12], 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. [1,3,5,6,9–11,13,15]).

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, under the assumptions that the GS method is well defined (in the sense that every subproblem has an optimal solution) and that the sequence $\{x^k\}$ admits limit points.

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 line search 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 every limit point of $\{x_k\}$ is a critical point of problem (1), even in the absence of any convexity assumption on f. As an example, we illustrate an application of the two-block GS method to the computation of critical points of nonconvex quadratic programming problems via the solution of a sequence of convex programming subproblems.

Then we consider the convergence properties of the GS method for the general case of a m-block decomposition under generalized convexity assumptions on the objective function. We show that the limit points of the sequence generated by the GS method are critical points of the constrained problem both when (i) f is componentwise strictly quasiconvex 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 [10,5]; in case (ii) we extend to the constrained case the results given in [14] for the cyclic coordinate method and in [6] for the unconstrained block GS method. Using a constrained version of a Powell's counterexample, we show also that nonconvergence of the GS method can be demonstrated for nonconvex functions, when $m \geqslant 3$ and the preceding assumptions are not satisfied.

Finally, in the general case of arbitrary decomposition, we extend a result of [1], by showing that the limit points of the sequence generated by a proximal point modification of the GS method are critical points of the constrained problem, 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 $x = (x_1, x_2, \ldots, x_m)$. In correspondence to this partition, the function value $f(x)$ is also indicated by $f(x_1, x_2, \ldots, x_m)$ and, for $i = 1, 2, \ldots, m$ the partial gradient of f with respect to x_i, evaluated at x, is indicated by $\nabla_i f(x) = \nabla_i f(x_1, x_2, \ldots, x_m) \in R^{n_i}$.

A *critical point* for Problem (1) is a point $\bar{x} \in X$ such that $\nabla f(\bar{x})^{\mathrm{T}}(y - \bar{x}) \geqslant 0$, for every $y \in X$, where $\nabla f(x) \in R^n$ denotes the gradient of f at x. If both $\bar{x}$ and y are partitioned into component vectors, it is easily seen that $\bar{x} \in X$ is a critical point for Problem (1) if and only if for all $i = 1, \ldots, m$ we have: $\nabla_i f(\bar{x})^{\mathrm{T}}(y_i - \bar{x}_i) \geqslant 0$ for every $y_i \in X_i$.

We denote by $\mathscr{L}_X^0$ the level set of f relative to X, corresponding to a given point $x^0 \in X$, that is $\mathscr{L}_X^0 := \{x \in X: f(x) \leqslant f(x^0)\}$. 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 line search 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 $z^k = (z_1^k, \ldots, z_m^k)$, with $z_i^k \in X_i$ for $i = 1, \ldots, m$. Let us choose an index $i \in \{1, \ldots, m\}$ and assume that for all k we can compute search directions

$$d_i^k = w_i^k - z_i^k \quad \text{with } w_i^k \in X_i, \tag{2}$$

such that the following assumption holds.

Assumption 1. Let $\{d_i^k\}$ be the sequence of search directions defined by (2). Then:
(i) there exists a number $M > 0$ such that $\|d_i^k\| \leqslant M$ for all k;
(ii) we have $\nabla_i f(z^k)^{\mathrm{T}} d_i^k < 0$ for all k.

An Armijo-type line search algorithm can be described as follows.

Line search algorithm (LS)
Data: $\gamma_i \in (0,1)$, $\delta_i \in (0,1)$.
Compute

$$\alpha_i^k = \max_{j=0,1,\ldots} \{(\delta_i)^j \colon f(z_1^k,\ldots,z_i^k + (\delta_i)^j d_i^k,\ldots,z_m^k)$$
$$\leqslant f(z^k) + \gamma_i(\delta_i)^j \nabla_i f(z^k)^{\mathrm{T}} d_i^k\}. \tag{3}$$

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 not be the result of a line search along d_i^k. However, this has no substantial effect in the convergence proof, which can be deduced easily from the known results (see e.g. [3]).

Proposition 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* (3);
(ii) *if $\{z^k\}$ converges to $\bar{z}$ and*:

$$\lim_{k\to\infty} f(z^k) - f(z_1^k,\ldots,z_i^k + \alpha_i^k d_i^k,\ldots,z_m^k)$$
$$= 0, \tag{4}$$

then we have

$$\lim_{k\to\infty} \nabla_i f(z^k)^T d_i^k = 0. \tag{5}$$

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:

3.1. GS Method

Step 0: Given $x^0 \in X$, set $k = 0$.
Step 1: For $i = 1,\ldots,m$ compute

$$x_i^{k+1} = \arg\min_{y_i \in X_i} f(x_1^{k+1},\ldots,y_i,\ldots,x_m^k). \tag{6}$$

Step 2: Set $x^{k+1} = (x_1^{k+1},\ldots,x_m^{k+1})$, $k = k+1$ and go to Step 1.

Unless otherwise specified, we assume in the sequel that the updating rule (6) is well defined, and hence that every subproblem has solutions.

We consider, for all k, the partial updates introduced by the GS method by defining the following vectors belonging to X:

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

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 (6) that $w(k,i)$ is the constrained global minimizer of f in the ith component subspace, and therefore it satisfies the necessary optimality condition:

$$\nabla_i f(w(k,i))^{\mathrm{T}}(y_i - x_i^{k+1}) \geqslant 0 \quad \text{for every } y_i \in X_i. \tag{7}$$

We can state the following propositions.

Proposition 2. *Suppose that for some $i \in \{0,\ldots,m\}$ 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

$$f(w(k+1,i)) \leqslant f(w(k,i)). \tag{8}$$

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 nonincreasing, 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

$$f(w(k+1,i)) \leqslant f(w(k+1,j)) \leqslant f(w(k,i))$$

$$\text{for } 0 \leqslant j \leqslant i,$$

and

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

Proposition 3. *Suppose that for some* $i \in \{1,\dots,m\}$ *the sequence* $\{w(k,i)\}$ *admits a limit point* $\bar{w}$. *Then we have*

$$\nabla_i f(\bar{w})^{\mathrm{T}}(y_i - \bar{w}_i) \geqslant 0 \quad \text{for every } y_i \in X_i \qquad (9)$$

and moreover

$$\nabla_{i*} f(\bar{w})^{\mathrm{T}}(y_{i*} - \bar{w}_{i*}) \geqslant 0 \quad \text{for every } y_{i*} \in X_{i*}, \quad (10)$$

where $i* = i \,(\mathrm{mod}\, m) + 1$.

Proof. Let $\{w(k,i)\}_K$ be a subsequence converging to $\bar{w}$. From (7), taking into account the continuity assumption on $\nabla_i f$, we get immediately (9).

In order to prove (10), 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

$$\nabla_{i+1} f(\bar{w})^{\mathrm{T}}(\tilde{y}_{i+1} - \bar{w}_{i+1}) < 0. \qquad (11)$$

Then, letting

$$d_{i+1}^k = \tilde{y}_{i+1} - w(k,i)_{i+1} = \tilde{y}_{i+1} - x_{i+1}^k$$

as $\{w(k,i)\}_K$ is convergent, we have that the sequence $\{d_{i+1}^k\}_K$ is bounded. Recalling (11) and taking into account the continuity assumption on $\nabla_i f$ it follows that there exists a subset $K_1 \subseteq K$ such that

$$\nabla_{i+1} f(w(k,i))^{\mathrm{T}} d_{i+1}^k < 0 \quad \text{for all } k \in K_1,$$

and therefore the sequences $\{w(k,i)\}_{K_1}$ and $\{d_{i+1}^k\}_{K_1}$ 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_{i+1}^k by means of Algorithm LS; then we have

$$f(x_1^{k+1},\dots,x_i^{k+1},x_{i+1}^k + \alpha_{i+1}^k d_{i+1}^k,\dots,x_m^k) \leqslant f(w(k,i)).$$

Moreover, as $x_{i+1}^k \in X_{i+1}$, $x_{i+1}^k + d_{i+1}^k \in X_{i+1}$, $\alpha_{i+1}^k \in (0,1]$, and X_{i+1} is convex, it follows that

$$x_{i+1}^k + \alpha_{i+1}^k d_{i+1}^k \in X_{i+1}.$$

Therefore, recalling that

$$f(w(k,i+1))$$
$$= \min_{y_{i+1} \in X_{i+1}} f(x_1^{k+1},\dots,x_i^{k+1},y_{i+1},\dots,x_m^k),$$

we can write

$$f(w(k,i+1))$$
$$\leqslant f(x_1^{k+1},\dots,x_i^{k+1},x_{i+1}^k + \alpha_{i+1}^k d_{i+1}^k,\dots,x_m^k)$$
$$\leqslant f(w(k,i)). \qquad (12)$$

By Proposition 2 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

$$\lim_{k\to\infty, k\in K_1} f(w(k,i)) - f(x_1^{k+1},\dots,x_i^{k+1},x_{i+1}^k$$
$$+ \alpha_{i+1}^k d_{i+1}^k,\dots,x_m^k) = 0.$$

Then, invoking Proposition 1, where we identify $\{z^k\}$ with $\{w(k,i)\}_{K_1}$, it follows that

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

which contradicts (11), so that we have proved that (10) holds when $i \in \{1,\dots,m-1\}$. When $i = m$, so that $i* = 1$, we can repeat the same reasonings noting that $w(k,m+1) = w(k+1,1)$. $\quad \square$

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 below.

Corollary 1. *Let* $\{x^k\}$ *be the sequence generated by the* GS *method and suppose that there exists a limit point* $\bar{x}$. *Then we have*

$$\nabla_1 f(\bar{x})^{\mathrm{T}}(y_1 - \bar{x}_1) \geqslant 0 \quad \text{for every } y_1 \in X_1 \qquad (13)$$

and

$$\nabla_m f(\bar{x})^{\mathrm{T}}(y_m - \bar{x}_m) \geqslant 0. \quad \text{for every } y_m \in X_m. \quad (14)$$

4. The two-block GS method

Let us consider the problem:

$$\text{minimize } f(x) = f(x_1, x_2),$$
$$x \in X_1 \times X_2 \qquad (15)$$

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 $(x_2, \ldots, x_N)$, then once x_1 is fixed, the objective function can be minimized in parallel with respect to the components x_i for $i = 2, \ldots, N$.

When $m=2$, a convergence proof for the GS method (2Block GS method) in the unconstrained case was given in [6]. Here the extension to the constrained case is an immediate consequence of Corollary 1.

Corollary 2. *Suppose that the sequence $\{x^k\}$ generated by the 2Block GS method has limit points. Then, every limit point $\bar{x}$ of $\{x^k\}$ is a critical point of Problem* (15).

As an application of the preceding result we consider the problem of determining a critical point of a nonlinear programming problem where the objective function is a nonconvex quadratic function and we have disjoint constraints on two different blocks of variables. In some of these cases the use of the two-block GS method may allow us to determine a critical point via the solution of a sequence of convex programming problems of a special structure and this may constitute a basic step in the context of cutting plane or branch and bound techniques for the computation of a global optimum.

As a first example, we consider a bilinear programming problem with disjoint constraints and we reobtain a slightly improved version of a result already established in [7] using different reasonings.

Consider a bilinear programming problem of the form:

minimize $f(x_1, x_2) = x_1^T Q x_2 + c_1^T x_1 + c_2^T x_2$

subject to $A_1 x_1 = b_1, \ x_1 \geqslant 0,$ (16)

$\quad\quad\quad\quad A_2 x_2 = b_2, \ x_2 \geqslant 0,$

where $x_1 \in R^{n_1}$ and $x_2 \in R^{n_2}$. As shown in [8], problems of this form can be obtained, for instance, as an equivalent reformulation on an extended space of concave quadratic programming problems.

Suppose that the following assumptions are satisfied:

(i) the sets $X_1 = \{x_1 \in R^{n_1}: A_1 x_1 = b_1, \ x_1 \geqslant 0\}$ and $X_2 = \{x_2 \in R^{n_2}: A_2 x_2 = b_2, \ x_2 \geqslant 0\}$ are non empty;

(ii) $f(x_1, x_2)$ is bounded below on $X = X_1 \times X_2$.

Note that we do not assume, as in [7] that X_1 and X_2 are bounded. Starting from a given point $(x_1^0, x_2^0) \in X$, the two-block GS method consists in solving a sequence of two linear programming subproblems. In fact, given (x_1^k, x_2^k), we first obtain a solution x_1^{k+1} of the problem

minimize $(Q x_2^k + c_1)^T x_1$

subject to $A_1 x_1 = b_1, \ x_1 \geqslant 0$ (17)

and then we solve for x_2^{k+1} the problem

minimize $(Q^T x_1^{k+1} + c_2)^T x_2$

subject to $A_2 x_2 = b_2, \ x_2 \geqslant 0.$ (18)

Under the assumption stated, it is easily seen that problems (17) and (18) have optimal solutions and hence that the two block GS method is well defined. In fact, reasoning by contradiction, assume that one subproblem, say (17), does not admit an optimal solution. As the feasible set X_1 is nonempty, this implies that the objective function is unbounded from below on X_1. Thus there exists a sequence of points $z^j \in X_1$ such that

$$\lim_{j \to \infty} (Q x_2^k + c_1)^T z^j = -\infty$$

and therefore, as x_2^k is fixed, we have also

$$\lim_{j \to \infty} f(z^j, x_2^k) = \lim_{j \to \infty} (Q x_2^k + c_1)^T z^j + c_2^T x_2^k = -\infty.$$

But this would contradict assumption (ii), since (z^j, x_2^k) is feasible for all j.

We can also assume that x_1^k and x_2^k are vertex solutions, so that the sequence $\{(x_1^k, x_2^k)\}$ remains in a finite set. Then, it follows from Corollary 2 that the two block-GS method must determine in a finite number of steps a critical point of problem (16).

As a second example, let us consider a (possibly nonconvex) problem of the form

minimize $f(x_1, x_2) = \frac{1}{2} x_1^T A x_1 + \frac{1}{2} x_2^T B x_2 + x_1^T Q x_2$
$\quad\quad\quad\quad\quad\quad + c_1^T x_1 + c_2^T x_2$

subject to $x_1 \in X_1,$

$\quad\quad\quad\quad x_2 \in X_2,$ (19)

where $X_1 \subseteq R^{n_1}, X_2 \subseteq R^{n_2}$ are nonempty closed convex sets, and the matrices A, B are symmetric and semidefinite positive.

Suppose that one of the following assumptions is verified:

 (i) X_1 and X_2 are bounded;

(ii) A is positive definite and X_2 is bounded.

Under either one of these assumptions, it is easily seen that the level set $\mathcal{L}_X^0$ is compact for every $(x_1^0, x_2^0) \in X_1 \times X_2$. This implies that the two-block GS method is well defined and that the sequence $\{x^k\}$ has limit points. Then, again by Corollary 2, we have that every limit point of this sequence is a critical point of problem (19).

5. The block GS method under generalized convexity assumptions

In this section we analyze 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 quasiconvexity of the objective function with respect to $m - 2$ components, or assuming that the objective function is pseudoconvex and has bounded level sets.

We state formally the notion of strict componentwise quasiconvexity that follows immediately from a known definition of strict quasiconvexity [10], which sometimes is called also strong quasiconvexity [2].

Definition 1. Let $i \in \{1, \dots, m\}$; we say that f is strictly quasiconvex with respect to $x_i \in X_i$ on X if for every $x \in X$ and $y_i \in X_i$ with $y_i \neq x_i$ we have

$$f(x_1, \dots, tx_i + (1 - t)y_i, \dots, x_m)$$
$$< \max\{f(x), f(x_1, \dots, y_i, \dots, x_m)\}$$
$$\text{for all } t \in (0, 1).$$

We can establish the following proposition, whose proof requires only minor adaptations of the arguments used, for instance, in [10,5].

Proposition 4. *Suppose that f is a strictly quasiconvex function with respect to $x_i \in X_i$ on X in the sense of Definition 1. 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:

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

Then, if $\lim_{k \to \infty} f(y^k) - f(v^k) = 0$, we have $\lim_{k \to \infty} \|v_i^k - y_i^k\| = 0$.

Then, we can state the following proposition.

Proposition 5. *Suppose that the function f is strictly quasiconvex with respect to x_i on X, for each $i = 1, \dots, m - 2$ in the sense of Definition 1 and that the sequence $\{x^k\}$ generated by the GS method has limit points. Then, every limit point $\bar{x}$ of $\{x^k\}$ is a critical point of Problem (1).*

Proof. Let us assume that there exists a subsequence $\{x^k\}_K$ converging to a point $\bar{x} \in X$. From Corollary 1 we get

$$\nabla_m f(\bar{x})^T (y_m - \bar{x}_m) \geq 0 \quad \text{for every } y_m \in X_m. \quad (20)$$

Recalling Proposition 2 we can write

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

Using the strict quasiconvexity assumption on f and invoking Proposition 4, 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 4 to the sequences $\{w(k, i - 1)\}_K$ and $\{w(k, i)\}_K$, for $i = 1, \dots, m - 2$, we obtain

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

Then, Proposition 3 implies

$$\nabla_i f(\bar{x})^T (y_i - \bar{x}_i) \geq 0 \quad \text{for every } y_i \in X_i,$$
$$i = 1, \dots, m - 1. \quad (21)$$

Hence, the assertion follows from (20) and (21). $\quad \square$

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

Proposition 6. *Suppose that f is pseudoconvex on X and that $\mathcal{L}_X^0$ is compact. Then, the sequence $\{x^k\}$ generated by the GS method has limit points and 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 $f(x^{k+1}) \leqslant f(w(k,i)) \leqslant f(x^k)$ for $i=0,\ldots,m$. Then, the points of the sequences $\{w(k,i)\}$, with $i = 0,\ldots,m$, belong to the compact set $\mathcal{L}_X^0$. Therefore, if $\bar{x} \in X$ is a limit point of $\{x^k\}$ we can construct a subsequence $\{x^k\}_K$ such that

$$\lim_{k\to\infty, k\in K} x^k = \bar{x} = \bar{w}^0, \tag{22}$$

$$\lim_{k\to\infty, k\in K} w(k,i) = \bar{w}^i \quad i=1,\ldots,m, \tag{23}$$

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

$$w(k,i) = w(k,i-1) + d(k,i) \quad \text{for } i=1,\ldots,m, \tag{24}$$

where the block components $d_h(k,i) \in R^{n_h}$ of the vector $d(k,i)$, with $h \in \{1,\ldots,m\}$, are such that $d_h(k,i) = 0$ if $h \neq i$. Therefore, for $i=1,\ldots,m$, from (22)–(24) we get

$$\bar{w}^i = \bar{w}^{i-1} + \bar{d}^i, \tag{25}$$

where

$$\bar{d}^i = \lim_{k\to\infty, k\in K} d(k,i) \tag{26}$$

and

$$\bar{d}_h^i = 0, \quad h \neq i. \tag{27}$$

By Proposition 2 we have

$$f(\bar{x}) = f(\bar{w}^i) \quad \text{for } i=1,\ldots,m. \tag{28}$$

From Proposition 3 it follows, for $i=1,\ldots,m$,

$$\nabla_i f(\bar{w}^i)^{\mathrm{T}}(y_i - \bar{w}_i^i) \geqslant 0 \quad \text{for all } y_i \in X_i, \tag{29}$$

and

$$\nabla_{i^*} f(\bar{w}^i)^{\mathrm{T}}(y_{i^*} - \bar{w}_{i^*}^i) \geqslant 0 \quad \text{for all } y_{i^*} \in X_{i^*}, \tag{30}$$

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

$$\nabla_\ell f(\bar{w}^j)^{\mathrm{T}}(y_\ell - \bar{w}_\ell^j) \geqslant 0 \quad \text{for all } y_\ell \in X_\ell, \tag{31}$$

then it follows

$$\nabla_\ell f(\bar{w}^{j-1})^{\mathrm{T}}(y_\ell - \bar{w}_\ell^{j-1}) \geqslant 0 \quad \text{for all } y_\ell \in X_\ell. \tag{32}$$

Obviously, (32) holds if $\ell = j$ (see (30)). Therefore, let us assume $\ell \neq j$. By (25)–(27) we have

$$\bar{w}^j = \bar{w}^{j-1} + \bar{d}^j, \tag{33}$$

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

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

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 (29) and (31), observing that (33) and the fact that $\ell \neq j$ imply

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

we obtain

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

It follows by the pseudoconvexity of f that

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

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

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

which, recalling the definition of $z(\eta)$, implies (32).

Finally, taking into account (30), and using the fact that (31) implies (32), by induction we obtain

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

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

As an example, let us consider a quadratic programming problem with disjoint constraints, of the form

$$\text{minimize} \quad f(x) = \sum_{i=1}^m \sum_{j=1}^m x_i^{\mathrm{T}} Q_{ij} x_j + \sum_{i=1}^m c_i^{\mathrm{T}} x_i \tag{34}$$

$$\text{subject to} \quad A_i x_i \geqslant b_i, \quad i=1,\ldots,m,$$

where we assume that:
(i) the sets $X_i = \{x_i \in R^{n_i}: A_i x_i \geqslant b_i\}$, for $i=1,\ldots,m$ are non empty and bounded;

(ii) the matrix

$$Q = \begin{pmatrix} Q_{11} & \dots & Q_{1m} \\ \dots & \dots & \\ Q_{m1} & \dots & Q_{mm} \end{pmatrix}$$

is symmetric and semidefinite positive.

In this case the function f is convex and the level set $\mathcal{L}_X^0$ is compact for every feasible x^0, but the objective function may be not componentwise strictly quasiconvex. In spite of this, the m-block GS method is well defined and, as a result of Proposition 6, we can assert that it converges to an optimal solution of the constrained problem, through the solution of a sequence of convex quadratic programming subproblems of smaller dimension.

6. A counterexample

In this section we consider a constrained version of a well-known counterexample due to Powell [12] 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 nonconvex function, which is componentwise convex but not strictly quasiconvex 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

$$\begin{aligned} f(x) = & -x_1x_2 - x_2x_3 - x_1x_3 + (x_1 - 1)_+^2 \\ & + (-x_1 - 1)_+^2 + (x_2 - 1)_+^2 + (-x_2 - 1)_+^2 \\ & + (x_3 - 1)_+^2 + (-x_3 - 1)_+^2, \end{aligned} \quad (35)$$

where

$$(t - c)_+^2 = \begin{cases} 0 & \text{if } t \leqslant c, \\ (t - c)^2 & \text{if } t \geqslant c. \end{cases}$$

Powell showed that, if the starting point x^0 is the point $(-1 - \varepsilon, 1 + \frac{1}{2}\varepsilon, -1 - \frac{1}{4}\varepsilon)$ 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 (35) 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 behavior evidenced by Powell is obtained if we consider a constrained problem with the same objective function (35) and a compact feasible set, defined by the box constraints

$$-M \leqslant x_i \leqslant M \quad i = 1, \dots, 3$$

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.

Nonconvergence 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 quasiconvex; 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 we can ensure convergence by requiring only the strict quasiconvexity 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. [1,3,4]), 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

Step 0: Set $k = 0$, $x^0 \in X$, $\tau_i > 0$ for $i = 1, \dots, m$..

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

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

$$+ \tfrac{1}{2} \tau_i \| y_i - x_i^k \|^2 \}. \quad (36)$$

Step 2: Set $x^{k+1} = (x_1^{k+1}, \dots, x_m^{k+1})$, $k = k + 1$ and go to Step 1.

The convergence properties of the method can be established by employing essentially the same arguments used in [1] in the convex cases and we can state the following proposition, whose proof is included here for completeness.

Proposition 7. *Suppose that the* PGS *method is well defined and that the sequence* $\{x^k\}$ *has limit points. Then every limit point* $\bar{x}$ *of* $\{x^k\}$ *is a critical point of Problem* (1).

Proof. Let us assume that there exists a subsequence $\{x^k\}_K$ converging to a point $\bar{x} \in X$. Define the vectors

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

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

Then we have

$$f(\tilde{w}(k, i))$$
$$\leqslant f(\tilde{w}(k, i-1)) - \tfrac{1}{2}\tau_i \|\tilde{w}(k, i) - \tilde{w}(k, i-1)\|^2,$$
$$(37)$$

from which it follows

$$f(x^{k+1})$$
$$\leqslant f(\tilde{w}(k, i)) \leqslant f(\tilde{w}(k, i-1)) \leqslant f(x^k)$$
$$\text{for } i = 1, \ldots, m. \tag{38}$$

Reasoning as in Proposition 2 we obtain

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

and hence, taking limits in (37) for $k \to \infty$ we have

$$\lim_{k\to\infty} \|\tilde{w}(k, i) - \tilde{w}(k, i-1)\| = 0, \quad i = 1, \ldots, m, \tag{39}$$

which implies

$$\lim_{k\to\infty, k\in K} \tilde{w}(k, i) = \bar{x}, \quad i = 0, \ldots, m. \tag{40}$$

Now, for every $j \in \{1, \ldots, m\}$, as x_j^{k+1} is generated according to rule (36), the point

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

satisfies the optimality condition

$$[\nabla_j f(\tilde{w}(k, j)) + \tau_j(\tilde{w}_j(k, j)$$
$$- \tilde{w}_j(k, j-1))]^{\mathrm{T}}(y_j - \tilde{w}_j(k, j)) \geqslant 0$$
$$\text{for all } y_j \in X_j.$$

Then, taking the limit for $k \to \infty$, $k \in K$, recalling (39), (40) and the continuity assumption on ∇f, for every $j \in \{1, \ldots, m\}$ we obtain

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

which proves our assertion. $\square$

Taking into account the results of Section 5, it follows that if the objective function f is strictly quasiconvex with respect to some component x_i, with $i \in \{1, \ldots, m\}$, then we can set $\tau_i = 0$. Moreover, reasoning as in the proof of Proposition 5, we can obtain the same convergence results if we set $\tau_{m-1} = \tau_m = 0$.

As an application of Proposition 7, let us consider the quadratic problem (34) of Section 5. Suppose again that the sets X_i are nonempty and bounded, but now assume that Q is an arbitrary symmetric matrix. Then the objective function will be not pseudoconvex, in general, and possibly not componentwise strictly quasiconvex. In this case, the block-GS method may not converge. However, the PGS method is well defined and, moreover, if we set $\tau_i > -2\lambda_{\min}(Q_{ii})$, then the subproblems are strictly convex and the sequence generated by the PGS method has limit points that are critical points of the original problem.

Acknowledgements

References

[1] A. Auslender, Asymptotic properties of the Fenchel dual functional and applications to decomposition problems, J. Optim. Theory Appl. 73 (3) (1992) 427–449.

[2] M.S. Bazaraa, H.D. Sherali, C.M. Shetty, in: Nonlinear Programming, Wiley, New York, 1993.

[3] D.P. Bertsekas, in: Nonlinear Programming, Athena Scientific, Belmont, MA, 1995.

[4] D.P. Bertsekas, P. Tseng, Partial proximal minimization algorithms for convex programming, SIAM J. Optim. 4 (3) (1994) 551–572.

[5] D.P. Bertsekas, J.N. Tsitsiklis, in: Parallel and Distributed Computation, Prentice-Hall, Englewood Cliffs, NJ, 1989.

[6] L. Grippo, M. Sciandrone, Globally convergent block-coordinate techniques for unconstrained optimization, Optim. Methods Software 10 (1999) 587–637.

[7] H. Konno, A cutting plane algorithm for solving bilinear programs, Math. Programming 11 (1976) 14–27.

[8] H. Konno, Maximization of a convex quadratic function under linear constraints, Math. Programming 11 (1976) 117–127.

[9] Z.Q. Luo, P. Tseng, On the convergence of the coordinate descent method for convex differentiable minimization, J. Optim. Theory Appl. 72 (1992) 7–35.

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

[11] M. Patriksson, Decomposition methods for differentiable optimization problems over Cartesian product sets, Comput. Optim. Appl. 9 (1) (1998) 5–42.

[12] M.J.D. Powell, On search directions for minimization algorithms, Math. Programming 4 (1973) 193–201.

[13] P. Tseng, Decomposition algorithm for convex differentiable minimization, J. Optim. Theory Appl. 70 (1991) 109–135.

[14] N. Zadeh, A note on the cyclic coordinate ascent method, Management Sci. 16 (1970) 642–644.

[15] W.I. Zangwill, in: Nonlinear Programming: A Unified Approach, Prentice-Hall, Englewood Cliffs, NJ, 1969.