Philip Wolfe

On the Convergence of Gradient Methods under Constraint

Abstract: The mathematical programming problem discussed is the convergence of a certain popular type of gradient procedure for maximizing a function under inequality constraints. An example shows that convergence to a solution need not always occur, and a theorem shows that under certain circumstances the gradient method does converge.

Introduction

A natural extension of Cauchy's method of steepest descent, presented in the next section, has often been used to solve the mathematical programming problem,

$$\text{Maximize } f(x_1, \cdots, x_m) \text{ subject to: } x_j \geq 0, \quad j = 1, \cdots, m. \tag{1}$$

See, for example, Refs. 1 and 2, which describe a procedure for reducing the general problem of maximizing a function under linear inequalities to a problem of the above form and solving that, which has worked very well in practice. Because of the simplicity of the extension, we long thought that convergence to a solution of the problem would be easy to prove under mild restrictions on the function f. This presumption turned out to be false: we eventually constructed an example of the form (1), given here in the third section, showing the method to have a real flaw, i.e., it could yield an infinite sequence of points converging to a point having nothing to do with the solution of the problem.

Others also have been misled by the plausibility of the method, and an erroneous proof of its effectiveness has been published [3]. As far as we know, only Zoutendijk anticipated the real difficulty, and he presented devices for circumventing it, as discussed here in the section on the counterexample. We find most of these devices unappealing, however, and further have not been able to construct a troublesome example for which f has bounded second derivatives, for the reasons indicated in the fourth section. We can thus hope that the procedure will be effective for problems in which f has suitable smoothness properties. As a first step in this direction, we then show that if f is quadratic, the difficulty cannot arise. This bare fact has no computational interest because there are much better ways of solving quadratic problems; but often in gradient procedures the convergence properties of the quadratic case can be shown to hold for the case of bounded second derivatives. Unfortunately we have not been able to do that, for, as demonstrated in the last section, the proof used for the quadratic case does not extend even to cubic functions. Thus, there remains a large gap in our knowledge of the behavior of the steepest descent method for constrained problems.

Steepest ascent with nonnegative variables

Cauchy's procedure for the unconstrained maximization of $f(x) = f(x_1, \cdots, x_m)$ generates the sequence $\{x^n\}$ as follows: Letting $g^n = \nabla f(x^n) = [\partial f(x)/\partial x_1, \cdots, \partial f(x)/\partial x_m]$, the point x^{n+1} is chosen to maximize f along the ray $\{x^n + tg^n; t \geq 0\}$. The extension below is the most natural we can think of when the constraints have the form $x \geq 0$: Given $x^n \geq 0$, we first modify the components of the direction g^n so that the ray $x^n + tg^n$ does not immediately leave the constraint set, and then choose x^{n+1} to maximize f on that part of the ray satisfying the constraints. We define

$$\delta_j(x) \equiv \begin{cases} 0 & \text{if } x_j = 0 \quad \text{and} \quad \partial f(x)/\partial x_j \leq 0 \\ 1 & \text{otherwise.} \end{cases} \tag{2}$$

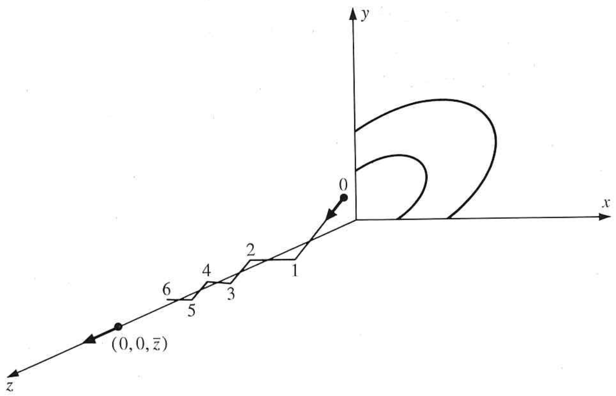

Figure 1 Convergence to a nonstationary point.

Thus $\delta_j(x) = 0$ precisely when motion in the direction of the gradient would cause x_j to become negative.

Beginning with $x^0 \geq 0$, the sequence $x^1, x^2, \cdots$ is obtained in this way:

Given x^n, define the vector g^n by

$$g_j^n \equiv \delta_j(x^n)\ \partial f(x)/\partial x_j . \tag{3}$$

Let

$$L_n = \sup\ \{t : x^n + tg^n \geq 0\}, \tag{4}$$

t_n be that value of t which maximizes $f(x^n + tg^n)$ for

$$0 \leq t \leq L_n, \tag{5}$$

$$\Delta x^n = t_n g^n \qquad \text{and} \qquad x^{n+1} = x^n + \Delta x^n . \tag{6}$$

In the absence of constraints we would set $\delta_j(x) = 1$, hence $g^n = \nabla f(x^n)$ and $L_n = +\infty$, reducing the step to Cauchy's.

We do not seem to get any results without assuming that ∇f is continuous, which is done henceforth. Curry [4] has shown that, in the absence of constraints, any point of accumulation $\bar{x}$ of the sequence $\{x^n\}$ is a stationary point of f, that is, that $\nabla f(\bar{x}) = 0$. (Curry actually defines t_n as the smallest value of t for which $f(x^n + tg^n)$ has a stationary point, but his theorem holds equally well when definition (5) is used.) In this constrained case we call $\bar{x}$ a stationary point when

$$\delta_j(\bar{x})\,(\partial f(\bar{x})/\partial x_j) = 0 \qquad \text{for all } j,$$

for precisely then do we have the necessary first-order conditions that $\bar{x}$ maximize f:

$$\nabla f(\bar{x}) \cdot \Delta x \leq 0 \qquad \text{for all } \Delta x \text{ such that } \bar{x} + \Delta x \geq 0 .$$

In order that some analogue of Curry's result will be valid in our case, we must at least know this: If $x^n \to \bar{x}$ as $n \to \infty$, then $\bar{x}$ is a stationary point. Given that, and some global information about f, such as its concavity, we could then show convergence to a global solution of the problem. See for example Ref. 2, Sect. 5. The example of the next section shows that the desired analogue does not hold.

Counterexample

Let $f(x, y, z) = -\frac{4}{3}(x^2 - xy + y^2)^{\frac{3}{4}} + z$. Letting $s = x^2 - xy + y^2$ we have

$$\nabla f(x, y, z) = [-s^{-\frac{1}{4}}(2x - y), -s^{-\frac{1}{4}}(-x + 2y), 1],$$

and since s behaves like $r^2 = x^2 + y^2$ $(\frac{1}{2}r^2 \leq s \leq \frac{3}{2}r^2)$, ∇f is continuous everywhere, but the second derivatives are not bounded in the neighborhood of $x, y = 0$. The function f is concave. Figure 1 sketches the level lines of f on the $z = 0$ plane and shows a few steps of the procedure, starting from a point $(0, a, z)$ such that $a \leq 2^{-1}$. As we show below, each step from one of the faces $x = 0$ or $y = 0$ terminates on the other face; the successive gradient directions tend toward that of the z-axis, but the steps taken shorten so quickly that the sequence of points tends to a limit point $(0, 0, \bar{z})$, which is not a constrained stationary point since the gradient there is $[0, 0, 1]$.

Let g^0 denote the gradient at $x^0 = (0, a, z)$; $g^0 = [a^{\frac{1}{2}}, -2a^{\frac{1}{2}}, 1]$. The ray $x^0 + tg^0$ pierces the other face for $t = \frac{1}{2}a^{\frac{1}{2}}$ in the point $x^1 = (\frac{1}{2}a, 0, z + \frac{1}{2}a^{\frac{1}{2}})$, where the gradient is $g^1 = [-(2a)^{\frac{1}{2}}, (\frac{1}{2}a)^{\frac{1}{2}}, 1]$. Since the rate of change of f along the ray at x^1 is positively proportional to $g^0 \cdot g^1 = 1 - 2(2a)^{\frac{1}{2}}$ the point x^1 will indeed be chosen by the procedure. (Since f is concave, its rate of change is monotone nonincreasing.) We see that

$$x^2 = [0, \tfrac{1}{4}a, b + \tfrac{1}{2}a^{\frac{1}{2}} + \tfrac{1}{2}(\tfrac{1}{2}a)^{\frac{1}{2}}],$$

$$x^3 = [\tfrac{1}{8}a, 0, b + \tfrac{1}{2}a^{\frac{1}{2}} + \tfrac{1}{2}(\tfrac{1}{2}a)^{\frac{1}{2}} + \tfrac{1}{2}(\tfrac{1}{4}a)^{\frac{1}{2}}]$$

and so forth, so that $x^n \to (0, 0, \bar{z})$, where

$$\bar{z} = z + \tfrac{1}{2}\sum_{j=0}^{\infty} (2^{-j}a)^{\frac{1}{2}} = z + (1 + \tfrac{1}{2}2^{\frac{1}{2}})\,a^{\frac{1}{2}} .$$

The example is not particularly "delicate." The exponent in the definition of the function might be any number between $\frac{1}{2}$ and 1, and some freedom in the choice of the function exponentiated is possible. As we shall see in the next section, however, it is not possible to make an example of this form having bounded second derivatives.

Zoutendijk conjectured the existence of the behavior exhibited by this example, calling it "zigzagging" in his study of a variety of procedures for nonlinear programming problems [5]. The simplest one for the present method (AZ1 of Zoutendijk) consists in altering definition (2) to

$$\delta_j(x) = \begin{cases} 0 & \text{if } x_j \leq \delta \text{ and } \partial f(x)/\partial x_j < 0, \\ 1 & \text{otherwise}, \end{cases}$$

and using this new δ_j in definition (3) of g^n. We can then show that if $x^n \to \bar{x}$ we have $\delta_j(\bar{x})\partial f(\bar{x})/\partial x_j = 0$ for all j. The price paid for thus weakening the notion of stationary point to obtain satisfactory convergence properties is this weakening of the first-order conditions that $\bar{x}$ maximize f; we can only assert that $\nabla f(x) \cdot \Delta x \leq 0$ for all Δx such that $\bar{x}_j + \Delta x_j \geq \delta$ for all j. Actually, for practical computation this seems to be a completely satisfactory result, especially since some such modification of the definition of δ_j would in any case have to be made to avoid ambiguities that are due to round-off error.

An interesting alternative not requiring the above weakening has been proposed by McCormick [6], who uses the modification he calls "bending": the ray $(x^n + tg^n : t \geq 0)$ of the ordinary method is replaced by the segment chain defined by

$$x_j^n(t) = \mathrm{Max}[0\,,\, x_j^n + t\,\partial f(x^n)/\partial x_j] \qquad \text{for each } j\,,$$

and x^{n+1} is given as $x^n(t^*)$, where t^* maximizes $f[x^n(t)]$ for $t \geq 0$. This appealing idea yields, for our example, a sequence of two-segment chains: the first segment from $(0\,, a\,, b)$ to $(\frac{1}{2}a\,, 0\,, b + \frac{1}{2}a^{\frac{1}{2}})$ as above, and the second segment from that point to $(\frac{1}{4}a\,, 0\,, b + \frac{1}{2}a^{\frac{1}{2}} + \frac{1}{4}a^{-\frac{3}{2}})$, which is thus the successor of $(0\,, a\,, b)$. The next point in the sequence evidently has the form $(0\,, a\,, b + \cdots)$ again, and the sequence moves out to infinity taking steps of two different sizes. That is not a happy outcome for an algorithm, of course, but nothing better can be expected of one so closely connected with steepest descent.

Consequences of convergence

The convergence of the series $\sum t_n$ in the example of the previous section plays an important part in a convergent descent process, as the lemma indicates.

Lemma

Let the operator H be bounded at zero; that is, $|H(x)| \leq K|x|$ for all sufficiently small x, and let the recursion $x^{n+1} = x^n + t_n H(x^n)$ be such that $x^n \to 0$ and $\sum |t_n| < \infty$. Then for some N, $x_N = x_{N+1} = \cdots = 0$.

Proof

$M_n = \mathrm{Max}\ (|x_k| : k \geq n)$ exists for all n and is monotone nonincreasing. Since $x^p - x^n - \sum_{k=n}^{p-1} \Delta x^k$, $x^n = -\sum_{k\geq n} \Delta x^k$ for all n, and

$$|x^n| \leq \sum_{k\geq n} |\Delta x^k| \leq K \sum_{k\geq n} |t_k|\ |x_k| \leq KM_n \sum_{k\geq n} |t_k|$$

$$\leq KM_N \sum_{k\geq N} |t_k| \qquad \text{for } n \geq N\,,$$

so that

$$M_N \leq M_N K \sum_{k\geq N} |t_k|\,.$$

Taking N so large that $K\sum_{k\geq N} |t_k| < 1$, the conclusion follows.

(If x^n is a sequence of numbers and $H(x^n) = -x^n$, then $x^{n+1} = (1 - t_n)x^n$, and the lemma is just the theorem that, if $\sum t_n$ is absolutely convergent, then the sequence of products $\prod_{k\geq n} (1 - t_n)$ has the limit zero only if some member vanishes.)

The lemma shows that trouble with second derivatives was an essential feature of the example of the previous section, whose function f had the "separated" form $h(x\,, y) + r(z)$. If the second derivatives of h were bounded and $\sum t_n$ converged, the lemma would apply to the $x\,, y$ steps alone, which would have to terminate. A nonterminating example for which f has bounded second derivatives cannot therefore have a "separated" form, which may explain why we have not been able to find such an example.

From now on we assume that the sequence x^n is given by the procedure (3) to (6) and converges to $\bar{x}$, stationary or not, but does not terminate.

For each component j there are three possibilities:

1) $x_j^n = 0$ for all sufficiently large n,
2) $x_j^n = 0$ for infinitely many n and $x_j^n > 0$ for infinitely many n,
3) $x_j^n > 0$ for all sufficiently large n.

Supposing the process to have begun sufficiently far out, the variables of type 1) may be discarded.

For those of type 2), evidently $\bar{x}_j = 0$, $\Delta x_j^n = t_n \delta_j(x^n)\ \partial f(x^n)/\partial x_j$, and the partial derivative is never eventually constant in sign. For those of type 3) it is always true that $\delta_j(x^n) = 1$ and $\Delta x_j^n = t_n\ \partial f(x^n)/\partial x_j$. Let us call the variables of type 2) y, and write $z_j^n = x_j^n - \bar{x}_j$ for all those of type 3). Substituting for the appropriate components of x, the recursion may then be written

$$\Delta y^n = T_n F(y^n\,, z^n)$$
$$\Delta z^n = t_n G(y^n\,, z^n)\,, \tag{8}$$

where T_n is a diagonal matrix formed from $t_n\delta_j(x^n)$ for appropriate j, and F and G are vector functions giving the appropriate partial derivatives of f. We have the following information about these equations:

Each component of F is infinitely often positive and negative,

$y^n\,, z^n \to 0$ as $n \to \infty$,

$y^n \geq 0\,, y^n \neq 0$ for infinitely many n.

Since F is continuous, $F(0\,, 0) = 0$, and that part gives no trouble about stationary points. Stationarity fails just when $G_j(0\,, 0) > 0$, or $G_j(0\,, 0) \lessdot 0$ while $x_j > 0$.

In any case if, say $G_j(0\,,0) = p_j \neq 0$, then choosing K so that $|p_j - G_j(y^k\,,z^k)| < \frac{1}{2}|p_j|$ for all $k \geq K$, we have for $n \geq K$

$$|z_j^{\,n} - z_j^{\,k}| = \left|\sum_{k=K}^{n-1} t_k G_j(y^k\,,z^k)\right| \geq \tfrac{1}{2}p_j \sum_{k=K}^{n-1} t_k\,,$$

so that $\sum t_n$ is convergent. Thus if $\bar{x}$ is not a stationary point, then $\sum t_n$ converges.

Quadratic case

If f is quadratic then F and G of the previous section are linear. The theorem below shows that in this case nonterminating convergence to a nonstationary point cannot occur. Note that, in the absence of the terms z^n, Δz^n, the theorem is a case of the lemma. The proof relies heavily on the linearity of F and G. Some remarks on the difficulty of extending the proof follow it.

• *Theorem*

Let $t_n > 0$, $\sum_{n=0} t_n < \infty$, and let the vector recursion

$$y^{n+1} = y^n + \Delta y^n\,,\ z^{n+1} = z^n + \Delta z^n$$

be specified by

$$\Delta y^n = T_n(Uy^n + Rz^n)\,, \tag{9}$$

$$\Delta z^n = t_n(p + Sy^n + Qz^n)\,, \tag{10}$$

and T_n, U, R, S, and Q are matrices of appropriate size. If: $\|T_n\| \leq t_n$ for all n; y^n, $z^n \to 0$ as $n \to \infty$; and no component of $Uy^n + Rz^n$ is eventually constant in sign, then for some N, $y^n = 0$ for all $n > N$.

Proof

In outline, the proof is as follows. If $p = 0$, the lemma can be immediately applied to the system, giving the result. If $p \neq 0$, then it appears that y^n must be much smaller than z^n, so that an approximation $\bar{z}^n$ to z^n can be found by setting $y^n = 0$ and solving the difference equation (10). The approximation $\bar{z}^n$ has a form similar to a power series in the variable $s_n = \sum_{k \geqslant n} t_k$, which tends to zero as $n \to \infty$. If the recursion relations are rewritten using the perturbation $w^n = z^n - \bar{z}^n$, then the term p naturally drops from (10) and the coefficients of the "power series" appear in (9); but, since Δy^n is never constant in sign, it can be argued one by one that the coefficients must vanish, and the resulting equations in y and w yield at once to the lemma.

The index N of the theorem is to be chosen large enough to satisfy several independent conditions: first, so that the classification 1) to 3) of the variables made in the previous section holds, and then so that the conditions (12), (14), (20), and (22) below will obtain. We suppose henceforth N to have been so chosen, and understand any formula involving the free subscript n to be asserted for all $n \geq N$ unless the contrary is stated.

Let $s_n^{(1)} = \sum_{k \geqslant n} t_n = s_n$ for all $n \geq 0$, and define recursively

$$s_n^{(i+1)} = \sum_{k \geqslant n} t_k s_k^{(i)} \qquad \text{for } i \geq 1\,.$$

Then $s_n^{(i)} \to 0$ as $n \to \infty$, each sequence $s_n^{(i)}$ is monotone nonincreasing, and for any $i > I > 0$,

$$s_n^{(i)} < s_n^{(I)}(s_n)^{i-I}\,. \tag{11}$$

Let N be such that

$$s_N\|Q\| < 1\,. \tag{12}$$

Let

$$\bar{z}^n = \sum_{i-1}^{\infty} (-1)^i\, s_n^{(i)}\, Q^{i-1}\, p\,,$$

which is well-defined by virtue of (11) and (12). It is easy to check that $\Delta\bar{z}^n = t_n(p + Q\bar{z}^n)$, so that, letting

$$z^n = w^n + \bar{z}^n\,,$$

the relations (9) and (10) become

$$\begin{aligned} \Delta y^n &= T_n\Big(Uy^n + Rw^n + \sum_{i \geqslant 1} (-1)^i\, s_n^{(i)}\, RQ^{i-1}\, p\Big) \\ \Delta w^n &= t_n(Sy^n + Qw^n)\,. \end{aligned} \tag{13}$$

The bulk of the proof consists in showing that $RQ^{i-1}p = 0$ for all i. Suppose that this is not the case, and that $I \geq 1$ is the first index i for which it is not.

Let $\operatorname{Max}_j\, |(RQ^{I-1}p)_j| = |(RQ^{I-1}p)_J| = q$.

Owing to (11) and (12) we may choose N so that

$$s_n^{(I)}q > 2 \sum_{i > 1} |(s_n^{(i)} RQ^{i-1}p)_j| \tag{14}$$

for all j, whence

$$\left|\Big[\sum_{i \geqslant 1} (-1)^i\, s_n^{(i)}\, RQ^{i-1}\, p\Big]_j\right| < \tfrac{3}{2}\, s_n^{(I)}\, q \tag{15}$$

for all j and

$$\tfrac{1}{2}\, s_n^{(I)}\, q < |[\text{(above)}]_J|\,. \tag{16}$$

Let the constant K be such that

$$|(Uy + Rw)_j|\,,\ |(Sy + Qw)_j| < K \operatorname{Max}\,(|y|\,,|w|) \tag{17}$$

for all y and w, and define

$$M_n = \operatorname*{Max}_{k \geqslant n} \operatorname{Max}\,(|y_k|\,,|w_k|)\,.$$

M_n exists for all $n \geq 0$, is monotone nonincreasing, and approaches zero as $n \to \infty$.

Now for any p, $n \geq 0$ we have $y^p - y^n = \sum_{k=n}^{p-1} \Delta y^k$. Letting $p \to \infty$,

$$y_n = -\sum_{k \geq n} \Delta y^k\,, \tag{18}$$

and similarly $w^n = -\sum_{k \geq n} \Delta w^k$. Thus

$$|y^n| \le \sum_{k\ge n} |\Delta y^k| ,$$

$$|w^n| \le \sum_{k\ge n} |\Delta w^k| .$$

Using successively these estimates, equations (13), the fact that $\|T_n\| \le t_n$, and the bounds (17) and (15), we find

$$\begin{aligned}|y^n|, |w^n| &\le \sum_{k\ge n} t_k \left[K \,\mathrm{Max}(|y^k|, |w^k|) + \tfrac{3}{2} s_k^{(I)} q\right] \\ &\le \sum_{k\ge n} t_k (KM_n + \tfrac{3}{2} s_k^{(I)} q) \\ &= Ks_nM_n + \tfrac{3}{2} s_n^{(I+1)} q .\end{aligned}$$

Since the right-hand side is monotone nonincreasing in n, for $k \ge n$ we have

$$|y^k|, |w^k| \le Ks_nM_n + \tfrac{3}{2} s_n^{(I+1)} q ,$$

whence $M_n \le Ks_nM_n + \frac{3}{2} s_n^{(I+1)} q$. Now suppose N is so large that

$$Ks_n < \tfrac{1}{2} . \tag{20}$$

Then $M_N \le 3 s_N^{(I+1)} q$, so that

$$|(Uy^n + Rw^n)_J| \le 3Ks_N^{(I+1)} q . \tag{21}$$

If finally N is chosen so that

$$3Ks_k^{(I+1)} q < \tfrac{1}{4} s_n^{(I)} q , \tag{22}$$

then (16) and (21) yield

$$\left|\left[Uy^n + Rw^n + \sum_{i\ge I} (-1)^i s_n^{(i)} RQ^{i-1} p - \tfrac{1}{2} s_n^{(I)} q\right]_J\right| < \tfrac{1}{4} s_n^{(I)} q ,$$

so that, for component J, the sign-changing property of the hypothesis of the theorem is contradicted. Thus in fact $RQ^{i-1}p = 0$ for all $i \ge 1$, and the relations (13) have the form

$$\Delta(y^n, w^n) = \bar{T}_n(y^n, w^n) \tag{23}$$

for a suitable matrix $\bar{T}_n$ such that $\|\bar{T}_n\| \le Kt_n$ for some suitable K. The lemma applies immediately to (23), so that $y^n = 0$ for n sufficiently large. The theorem is proved.

Nonquadratic case

The argument used in the proof above cannot be applied when the functions of (9) and (10) are not linear; we have stated the theorem in such a way that a counter-example can be given for the case in which (9) is quadratic. The counter-example presented here does not, however, actually show that the procedure (3) to (6) could not solve such a problem, because the functions are not in fact obtained from a gradient, nor do we know that the sequence t_n we use might arise from a gradient process.

Let y have one component and z two components, and set $F(y, z) = Rz + zWz$ and $G(y, z) = p + Qz$, with W and Q square and symmetric. Equations (9) and (10) are then replaced by

$$\Delta y^n = t_n(Rz^n + z^nWz^n) , \; \Delta z^n = t_n(p + Qz^n) .$$

The system can be solved for z^n and, as before for $\bar{z}^n$,

$$z^n = \sum_{i\ge 1} (-1)^i s_n^{(i)} Q^{i-1} p .$$

Substituting and arranging by terms $s_n^{(i)}$ of similar orders of magnitude, we have

$$Rz^n + z^nWz^n = -s_nRp + s_n^{(2)}RQp + s_n^{2}pWp + \mathrm{O}(s_n^3) .$$

An easy argument on orders of magnitude of s_n yields that $Rp = 0$ if the above expression changes sign infinitely often, but the argument cannot be pursued through the next two terms to show that they vanish, because they have comparable magnitudes. This seems to be exactly the typical difficulty encountered in attempting to extend the result from the quadratic case. To see that the difficulty is serious, let

$$p = \begin{pmatrix}1\\1\end{pmatrix}, \; Q = \begin{pmatrix}\frac{1}{2} & 0\\ 0 & 1\end{pmatrix}, \; R = (1\,,-1) \text{ , and } W = \frac{3}{16}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}.$$

Then $Rz^n + z^nWz^n = -\frac{1}{2}s_n^{(2)} + \frac{3}{8}s_n^2 + \mathrm{O}(s_n^3)$. Let the sequence $t_0, t_1, \cdots$ be $1, 1, 0.1, 0.1, 0.01, 0.01, \cdots$, and let $r_n = -\frac{1}{2}s_n^{(2)} + \frac{3}{8}s_n^2$. It is easy to calculate that

$$r_{2m} = \frac{100}{891}(10)^{-m} \quad \text{and} \quad r_{2m+1} = -\frac{487}{7128}(10)^{-m}$$

for $m = 0, 1, \cdots$, showing that $Rz^n = z^nWz^n$ eventually alternates in sign with increasing n. Thus this example satisfies the hypotheses of the theorem, excepting that F is quadratic rather than linear, but y^n does not eventually vanish.

References

1. P. Wolfe, "Methods of Nonlinear Programming," *Recent Advances in Mathematical Programming*, (R. L. Graves and P. Wolfe, eds.), McGraw-Hill Book Co., 1963, pp. 67–86, Section 6.
2. P. Faure and P. Huard, "Résolution de Programmes Mathématiques à Fonction Non Linéaire par la Méthode du Gradient Réduit," *Revue Française de Recherche Opérationelle*, No. 36, 167–206 (1965).
3. G. Hadley, *Nonlinear and Dynamic Programming*, Addison-Wesley Publishing Co., Inc., 1964. (The discussion referenced is in Sec. 9.3. It goes wrong in the third and fourth lines of p. 304.)
4. H. B. Curry, "The Method of Steepest Descent for Non-linear Minimization Problems," *Quarterly for Applied Mathematics* **2,** 258–261 (1944).
5. G. Zoutendijk, *Methods of Feasible Directions*, Elsevier Publishing Co., Amsterdam, 1960, Ch. 7.
6. Garth P. McCormick, "Anti-zig-zagging by bending" *Management Science* **15,** 315–320 (1969).

Received September 27, 1971

The author is located at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598.