University of Maryland College Park

Institute for Advanced Computer Studies TR–95–92
Department of Computer Science TR–3534

On Sublinear Convergence*

G. W. Stewart†

October 1995

ABSTRACT

This note develops a theory of sublinearly converging sequences, including a categorization of the rates of convergence and a method for determining the rate from an iteration function.

*This report is available by anonymous ftp from `thales.cs.umd.edu` in the directory `pub/reports`.

†Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742. This work was supported in part by the National Science Foundation under grant CCR 95503126.

On Sublinear Convergence

G. W. Stewart

ABSTRACT

This note develops a theory of sublinearly converging sequences, including a categorization of the rates of convergence and a method for determining the rate from an iteration function.

A sequence $x_1, x_2, \ldots$ converges sublinearly to x_* if

$$\lim_{x \to \infty} \frac{|x_{k+1} - x_*|}{|x_k - x_*|} = 1.$$

Alternatively, if $e_k = |x_k - x_*|$ denotes the error in x_k as an approximation to x_*, then

$$e_{k+1} = \lambda_k e_{k-1},$$

where $\lambda_k \to 1$. Thus sublinear convergence is slow convergence in the sense that the step-by-step reduction in the error becomes arbitrarily small with increasing k. For this reason algorithms that converge sublinearly are not greatly in evidence, and there is no general theory of sublinear convergence.

Sublinearly converging sequences, however, occur in contexts other than iterative algorithms. For example, the inspiration for this note was a recursion of the form

$$\sigma_{k+1} = \frac{\beta_k \sigma_k}{\sqrt{\beta_k^2 + \sigma_k^2}}, \tag{1}$$

in which the σ_k are lower bounds on the singular values of a sequence of matrices. For constant β_k, say $\beta_k = \beta$, the sequence converges sublinearly to zero — a good thing in the context, since the object of the investigation was to show that the singular values remain near one [2]. But a complete theory requires a knowledge of how quickly the bounds approach zero. From a numerical example I was able to deduce that for constant β

$$\sigma_k = \frac{1}{\sqrt{k}}$$

is a solution of (1). However, the fact that this deduction was largely accidental shows the need for a general method for determining the asymptotic behavior of a sublinearly converging sequence generated by a formula like (1). The purpose of this note is to provide such a method.

We will begin with a definition.

Definition 1. *Let* $\lim_{k\to\infty} x_k = x_*$, *and let* $e_k = |x_k - x_*|$. *Then* x_k *converges sublinearly with degree* $p > 0$ *if*

$$\lim_{k\to\infty} \frac{e_k - e_{k+1}}{e_k^{1+\frac{1}{p}}} = c > 0.$$

The constant c *is called the convergence constant.*

Alternatively, the convergence is sublinear of degree p if

$$e_{k+1} = e_k(1 - c_k e_k^{1/p}), \qquad c_k \to c > 0. \tag{2}$$

In this form it is easy to see that the error is reduced by a factor approaching one. The smaller p, the closer the factor to one. Thus small p means slow convergence.

The prototypical pth degree sequence is $x_k = k^{-p}$. To see this, expand $(k+1)^{-p}$ in a Taylor series with remainder to get

$$(k+1)^{-p} = k^{-p}(1 + k^{-1})^{-p} = k^{-p}\{1 - [p + o(1)]^{-1}k^{-1}\}.$$

Thus the errors $e_k = k^{-p}$ satisfy (2) with $c = 1/p$.

We are going to show that all convergence of degree p is essentially the same. However, first let us establish some facts about the errors of such a sequence.

In the first place, since

$$e_{k+1} = e_0 \prod_{i=0}^{k} (1 - c_i e_i^{1/p})$$

we must have

$$\prod_{i=0}^{\infty} (1 - c_i e_i^{1/p}) = 0.$$

Now it is well known that for $a_i > 0$

$$\prod_{i=0}^{\infty} (1 + a_i) = \infty \quad \text{or} \quad \prod_{i=0}^{\infty} (1 - a_i) = 0 \tag{3}$$

if and only if

$$\sum_{i=0}^{\infty} a_i = \infty.$$

It follows that

$$\sum_{i=0}^{\infty} e_i^{1/p} = \infty. \tag{4}$$

We are now in a position to establish the essential identity of all degree p convergence.

Theorem 2. *Let x_k and $\hat{x}_k$ both converge sublinearly with degree p and convergence constants c and $\hat{c}$. Let e_k and $\hat{e}_k$ denote their errors. Then*

$$\rho_k \equiv \frac{e_k}{\hat{e}_k} \to \frac{\hat{c}}{c}.$$

Proof. Let

$$e_{k+1} = e_k(1 - c_k e_k^{1/p}), \qquad c_k \to c > 0$$

and

$$\hat{e}_{k+1} = \hat{e}_k(1 - \hat{c}_k \hat{e}_k^{1/p}), \qquad \hat{c}_k \to \hat{c} > 0.$$

Then

$$\rho_{k+1} = \rho_k \frac{1 - c_k e_k^{1/p}}{1 - \hat{c}_k \hat{e}_k^{1/p}} = \rho_k \frac{1 - c_k \rho_k^{1/p} \hat{e}_k^{1/p}}{1 - \hat{c}_k \hat{e}_k^{1/p}}.$$

Expanding the denominator in a Neumann series and absorbing the error term in $\hat{c}_k$, we get

$$\rho_{k+1} = \rho_k[1 - (c_k \rho_k^{1/p} - \tilde{c}_k)\hat{e}_k^{1/p}], \tag{5}$$

where $\tilde{c}_k \to \hat{c}$.

Now $\rho_k^{1/p}$ cannot be uniformly greater than $\hat{c}/c$. For otherwise $(c_k \rho_k^{1/p} - \tilde{c}_k)$ is in the limit uniformly positive, and

$$\sum_{k=1}^{\infty} (c_k \rho_k^{1/p} - \tilde{c}_k)\hat{e}_k^{1/p} = \infty$$

[cf. (4)]. It follows from (3) and (5) that $\rho_k \to 0$, a contradiction. A similar argument shows that $\rho_k^{1/p}$ cannot be uniformly less than $\hat{c}/c$. It follows that for any $\epsilon > 0$, there are arbitrarily large k such that $|\rho_k - \hat{c}/c| < \epsilon$

Now let $\epsilon > 0$. Since $\hat{e}_k \to 0$, $c_k \to c$, and $\tilde{c}_k \to \hat{c}$, there is an ℓ such that for all $k \geq \ell$

$$\begin{array}{ll} 1. & |\tilde{c}_k/c_k - \hat{c}/c| < \epsilon/2, \\ 2. & |\rho_k - \hat{c}/c| < \epsilon \implies |\rho_{k+1} - \rho_k| < \epsilon/2. \end{array} \tag{6}$$

Let $k \geq \ell$ be such that $|\rho_k - \hat{c}/c| < \epsilon$. Our theorem will be proved if we can show that $|\rho_{k+1} - \hat{c}/c| < \epsilon$. There are three cases.

1. $|\rho_k - \hat{c}/c| < \epsilon/2$. In this case (6-1) implies that $|\rho_{k+1} - \hat{c}/c| < \epsilon$.

2. $\rho_k \geq \hat{c}/c + \epsilon/2$. By (6-2), $\rho_k > \hat{c}_k/c_k$, and it follows from (5) that $\rho_{k+1} < \rho_k$. Hence by (6-2), $|\rho_{k+1} - \hat{c}/c| < \epsilon$.

3. $\rho_k \leq \hat{c}/c - \epsilon/2$. By (6-2), $\rho_k < \hat{c}_k/c_k$, and it follows from (5) that $\rho_{k+1} > \rho_k$. Hence by (6-2), $|\rho_{k+1} - \hat{c}/c| < \epsilon$. ■

It follows from this theorem that if x_k converges to x_* with degree p and constant c, then

$$|x_k - x_*| \cong \frac{1}{pc}\frac{1}{k^p}. \tag{7}$$

Since for $\hat{p} > p$, $k^{-p} = o(k^{-\hat{p}})$, we have the following corollary.

Corollary 3. *If x_k and $\hat{x}_k$ converge with degrees p and $\hat{p} > p$, then*

$$e_k = o(\hat{e}_k).$$

Let φ be a function with a fixed point x_*. A classical result (e.g., see [1, p. 64]) says that the iteration

$$x_{k+1} = \varphi(x_k), \qquad k = 0, 1, \ldots$$

converges locally to x_* provided $|\varphi'(x_*)| < 1$. Moreover if

$$0 = \varphi'(x_*) = \cdots = \varphi^{(p-1)}(x_*) \neq \varphi^{(p)}(x_*),$$

then the convergence is superlinear of order p. If $|\varphi'(x_*)| = 1$ the convergence, if any, is sublinear, and we can determine its degree by inspecting derivatives.

Theorem 4. *Let $\varphi(x_*) = x_*$. Suppose that the qth derivative of φ is continuous at x_* and*

$$\varphi'(x_*) = 1 \quad \text{and} \quad 0 = \varphi''(x_*) = \cdots = \varphi^{(q-1)}(x_*) \neq \varphi^{(q)}(x_*).$$

Then for all x_0 sufficiently near x_* that satisfy

$$(x_0 - x_*)^{q-1}\varphi^{(p)}(x_*) < 0, \tag{8}$$

The iteration

$$x_{k+1} = \varphi(x_k), \qquad k = 0, 1, \ldots$$

converges sublinearly and monotonically to x_* with degree $p = 1/(q-1)$ and constant $|\varphi^{(p)}(x_*)|/q!$.

Proof. From a Taylor series expansion about x_* we have

$$\varphi(x) - x_* = (x - x_*) + \frac{\varphi^{(p)}(\xi)}{q!}(x - x_*)^q,$$

where ξ lies between x_* and x. Choose x_0 (if possible) to satisfy (8) and also so that

$$\frac{\varphi^{(p)}(\xi)}{q!}(x - x_*)^{q-1} < 1$$

for x between x_* and x_0. Then

$$x_{k+1} - x_* = (x_k - x_*)\left(1 - \frac{\varphi^{(p)}(\xi)}{q!}(x_k - x_*)^{q-1}\right)$$

For $k = 0$ the second factor in this relation is positive and less than one. Consequently, x_1 is strictly between x_* and x_0. An obvious induction will establish that the x_k proceed monotonically toward x_*. Moreover, for $x_k - x_*$ bounded away from zero, the decrease is uniform. Hence $x_k \to x_*$. The facts about the sublinear convergence of the x_k follow directly from the definition. ■

The regions from which the iteration converges depends on q. If $q - 1$ is odd, the iteration converges from the left or the right, depending on whether $\varphi^{(q)}(x_*)$ is positive or negative. If $q - 1$ is even there is no local convergence when $\varphi^{(q)}(x_*)$ is positive, and the convergence is from either side when $\varphi^{(q)}(x_*)$ is negative.

There is an analogous theorem for $\varphi'(x_*) = -1$; however, the situations under which convergence occurs is even more restricted. Specifically $q - 1$ must be even and $\varphi^{(q)}(x_*)$ must be positive. In this case, the errors converge monotonically to zero, but the x_k alternate about x_*.

It should be clear from the above discussion that the theory — unlike the theory of linear and higher order convergence — does not extend to functions of a complex variable.

Nor does the theory include all kinds of sublinear convergence. For example, the sequence $\frac{1}{k \log k}$ converges sublinearly to zero, but it lies between convergence of degree one and $1 + \epsilon$ for any $\epsilon > 0$. Such intermediate modes of convergence also appear in the theory of higher order convergence.

It is a curiosity of Theorem 4, that no analytic iteration function φ can produce sublinear convergence of degree greater than one.

Returning to the example that motivated these investigations, let

$$\varphi(\sigma) = \frac{\beta\sigma}{\sqrt{\beta^2 + \sigma^2}}.$$

Then it is easy to verify that

$$\varphi'(0) = 1, \quad \varphi''(0) = 0, \quad \varphi'''(0) = -\frac{3}{\beta}.$$

It follows from Theorem 4, that convergence is of degree $\frac{1}{2}$ with constant $\frac{1}{2\beta}$. Moreover, from (7), the iterates generated by φ must have the asymptotic form

$$\sigma_k \cong \frac{\beta}{\sqrt{k}}.$$

References

[1] K. E. Atkinson. *An Introduction to Numerical Analysis.* John Wiley, New York, 1978.

[2] G. W. Stwart. The triangular matrices of Gaussian elimination and related decompositions. Technical Report CS-TR-3533 UMIACS-TR-95-91, University of Maryland, Department of Computer Science, 1995.