

**NOTICE WARNING CONCERNING COPYRIGHT RESTRICTIONS:**  
The copyright law of the United States (title 17, U.S. Code) governs the making of photocopies or other reproductions of copyrighted material. Any copying of this document without permission of its author may be prohibited by law.

# Optimal Order and Minimal Complexity of One-Step Methods for Initial Value Problems

Arthur G. Werschulz

September, 1976

Department of Mathematics  
Carnegie-Mellon University  
Pittsburgh, PA 15213

Abstract: We consider the task of numerically approximating the solution of an ordinary differential equation initial value problem. A methodology is given for determining the computational complexity of finding an approximate solution with error not exceeding  $\epsilon$ . In addition, we determine the method of optimal order within a given class of methods, and show that under reasonable hypotheses, the optimal order increases as  $\epsilon$  decreases, tending to infinity as  $\epsilon$  tends to zero.

This research was supported in part by the National Science Foundation under Grant MCS75-222-55 and the Office of Naval Research under Contract N00014-76-C-0370, NR 044-422.

## 1. Introduction

With few exceptions, past work in analytic computational complexity has focused on the problem of finding a zero of a (nonlinear) transformation of Banach spaces; in most work, this problem is specialized to that of finding a zero of an operator on a finite-dimensional real or complex vector space (and in much of this work, the problem is further specialized to the one-dimensional case). Much has been discovered about the computational aspects of iterative schemes for the solution of such problems, especially in the areas of minimal complexity (e.g., Kung and Traub [73], Traub and Woźniakowski [76]) and maximal order (e.g., Kung and Traub [74], Woźniakowski [75]).

In this paper, we will consider another topic in analytic computational complexity theory, that of finding complexity bounds for the numerical solution of ordinary differential equation initial-value problems on a fixed interval. We will not be interested in questions of the existence and the uniqueness of the solutions to such problems. In fact, we will restrict our discussion of the application of general results to the case where the unique solutions to these problems are analytic functions. (The techniques described in this paper are applied to several well-known classes of methods in Werschulz [76a], [76b].)

We will limit ourselves here to classes of one-step methods for the numerical solution of these problems; in terms of informational usage, these methods are analogous to iterative zero-finding methods without memory (Traub [64], [72]). Analogous to the one-point iterative methods with memory are the multistep methods for initial-value problems; these methods will be dealt with in a future paper.

Our approach will be to assume that an initial-value problem is given, along with

some error criterion  $\epsilon$ , where  $0 < \epsilon < 1$ ; we then wish to compute an approximate solution with error no greater than  $\epsilon$ . Two basic questions concern us:

- (1.) For any given method, what is the complexity of solving this problem?
- (2.) Given any "basic" sequence of methods with increasing order, which method has minimal complexity?

In Section 2, we describe a methodology that handles both questions for classes consisting of methods whose error functions have a special form. Furthermore, we find that within such a basic sequence of methods, the following hold under very general conditions:

- (1.) For any  $\epsilon$ , there is a unique choice of order and step size minimizing the complexity.
- (2.) As  $\epsilon$  decreases, both the optimal order and the complexity increase monotonically, tending to infinity as  $\epsilon$  tends to zero.

Furthermore, within many classes of problems and methods, the "penalty" (e.g., the amount the cost curve turns near the optimum) associated with using non-optimal order tends to infinity as  $\epsilon$  tends to zero.

These conclusions are an interesting contrast to known results on zero-finding via iterations without memory. The latter results tend to support the "folklore" idea that it is "better" to use a low-order method many times, than to use a high-order method a few times. In the one-point case, optimal order is low, while in the multipoint case, optimal order increases with the problem complexity (but with little penalty for using a method of non-optimal order) (Kung and Traub [73]). In addition, optimal order for these problems does not depend on the error criterion; it is computed for the limiting case as  $\epsilon$  approaches zero.

One may wonder why there is this discrepancy between the results for the initial-value problem and those for the zero-finding problem, since any initial-value

problem may be written as an operator equation, as in Stetter [73]. The reason for this is that the methods used for the two problems differ greatly--those for the initial-value problem compute estimates for solution values at new points by discretization, while those for zero-finding compute improved estimates for the zero of a function by iteration.

In Section 3, we discuss the extension of these results to classes consisting of methods whose error functions are somewhat more complicated than those considered in Section 2.

In Section 4, we introduce the notions of normality and order-convergence for a basic sequence of one-step methods. We prove that they are equivalent under certain circumstances. A basic sequence of methods enjoying these properties is very easy to deal with in many respects, especially when one is interested in comparing upper and lower complexity bounds for such a class.

Finally, in Section 5, we describe some numerical data that support the above theoretical results. In particular, these data seem to indicate that even for modest values of  $\epsilon$ , there are considerable savings in using methods of optimal, rather than fixed, order.

## 2. Optimality Within a Strong Basic Sequence

We are interested in the numerical solution of a class of ordinary differential equation initial-value problems on a fixed interval  $I$  of finite length; we take  $I = [0, 1]$  without loss of generality. More precisely, let  $\mathcal{D}$  be a set of initial-value points in the real  $N$ -dimensional linear space  $\mathbb{R}^N$ , and let  $\mathcal{V}$  be a set of operators on  $\mathbb{R}^N$ , such that the initial-value problem of finding a function  $x: I \rightarrow \mathbb{R}^N$  satisfying

$$(2.1) \quad \begin{aligned} \dot{x}(t) &= v(x(t)) \quad \text{if } t \in \text{int } I, \\ x(0) &= x_0 \end{aligned}$$

has a unique solution for every  $(x_0, v) \in \mathcal{D} \times \mathcal{V}$ . The autonomous form of this system is no restriction, since any non-autonomous system may be made autonomous by increasing the dimension of the system by one.

The model of computation to be used is fairly general. We assume only that all arithmetic operations are performed exactly in  $\mathbb{R}$  (i.e., infinite-precision arithmetic), and that for any algorithm to be considered for the solution of (2.1), a set of procedures is given for the computation of any information about  $v$  required by that algorithm. (For instance, with Runge-Kutta methods, we must be able to compute  $v$  at any point in its domain.)

In this paper, we are interested in the numerical solution of (2.1) via one-step methods, using an equidistant grid as defined in Stetter [73]. (We limit ourselves to equidistant grids in order to facilitate the comparison of methods of different orders; the other extreme is taken by Lindberg [74], who considers the problem of picking an optimal grid for a given method of fixed order.) Thus the methods considered will generate approximations  $x_i$  to  $x(t_i)$  by the recursion

$$(2.2) \quad x_{i+1} = x_i + h \varphi(x_i, h) \quad (0 \leq i \leq n-1, n = h^{-1}),$$

where  $h$  is the step-size and  $\varphi$  is the increment function for the method (Henrici [62]); for brevity, we will refer to "the method  $\varphi$ ." Despite the fact that  $\varphi(x_i, h)$  will depend on some information about  $v$ , we will not explicitly indicate this dependence.

Thus, the method  $\varphi$  produces an approximation to the true solution of (2.1). We want to measure the discrepancy between the approximate and true solutions. Various error measures have been introduced in the literature. These include the local truncation error per step, the local truncation error per unit step, and the global error; see Henrici [62] or Stetter [73] for definitions. These error measures may be either absolute or relative (in the usual sense); they may be measured either at the endpoint of the interval (as in Henrici [62], Hindmarsh [74]) or over the entire grid (as in Sandberg [67], Lindberg [74]). There has been a great deal of discussion of which error criterion is the best one to use; for instance, Gear [71] (Section 9.3) uses local error per step, while Hull et al. [72] use local error per unit step. We take no sides in this discussion, since any of these error measures may be used in the analysis to follow.

Before proceeding any further, we will establish some notational conventions. Let  $\mathfrak{X}$  be an ordered ring; then  $\mathfrak{X}^+$  and  $\mathfrak{X}^{++}$  will respectively denote the nonnegative and positive elements of  $\mathfrak{X}$ . (This will be used in the cases  $\mathfrak{X} = \mathbb{R}$ , the real numbers, and  $\mathfrak{X} = \mathbb{Z}$ , the integers.) The symbol ":= " means "is defined to be," while "≡" means "is identically equal to." If  $x_1, x_2: \mathbb{R} \rightarrow \mathbb{R}$  and  $\omega: \mathbb{R}^2 \rightarrow \mathbb{R}$  are differentiable, then for  $i = 1, 2$ , we will write

$$\partial_i \omega(x_1(t), x_2(t))$$

for the result of differentiating  $\omega(x_1, x_2)$  with respect to  $x_i$ , and then substituting  $x_1 = x_1(t)$ ,  $x_2 = x_2(t)$ . We use the notations " $x \downarrow a$ " and " $x \uparrow a$ " to indicate one-sided

limits as in Buck [65]. Finally, we shall write "(a.b)<sub>c</sub>" to indicate the c<sup>th</sup> part of equation (a.b), as in Gurtin [75].

Now we are prepared to define our problem. Let  $\mathcal{D}$  and  $\mathcal{V}$  be as above; consider a problem  $(x_0, v)$  in  $\mathcal{D} \times \mathcal{V}$ . Let  $\Phi$  be a class of one-step methods, and let  $\sigma: \Phi \times I \rightarrow \mathbb{R}^+$  satisfying  $\lim_{h \downarrow 0} \sigma(\varphi, h) = 0$  be a given function that will serve as an error measure. Choose an error criterion  $\epsilon$  satisfying the technical restriction  $0 < \epsilon < 1$ .

1. We then wish to answer two questions:

(1.) Given  $\varphi \in \Phi$ , how may we pick  $h \in I$  such that

$$(2.3) \quad \sigma(\varphi, h) \leq \epsilon,$$

and what is the complexity of the process defined by  $\varphi$  and  $h$ ?

(2.) How may one choose among all  $(\varphi, h) \in \Phi \times I$  such that (2.3) holds, that pair  $(\varphi^*, h^*)$  giving minimal complexity?

In order to get useful bounds on  $\sigma(\varphi, h)$ , it is necessary to introduce the concept of order. In this section, we will use a highly restricted definition, which we will relax in Section 3. Let  $\Phi = \{\varphi_p; p \in \mathbb{Z}^{++}\}$ , and suppose that there is an analytic function  $\kappa: \mathbb{R}^+ \rightarrow \mathbb{R}^+$  such that  $\lim_{p \rightarrow 0} \kappa(p)^{1/p}$  exists and is nonzero and

$$(2.4) \quad \sigma(\varphi_p, h) = \kappa(p) h^p \quad \text{for } h \in I \text{ and } p \in \mathbb{Z}^{++}.$$

Then  $\varphi_p$  is said to have strong order  $p$  with respect to  $\sigma$ , and  $\Phi$  is said to be a strong basic sequence. (Although the error coefficient  $\kappa$  will generally depend on the solution  $x$  of (2.1), we do not explicitly indicate this dependence.) Note that the order of a method depends on the error measure; for example, the order with respect to the local error per step is one greater than that with respect to the local error per unit step or the global error.

Equation (2.4) is somewhat more restrictive than that which is usually encountered in practice; more often, we expect  $\kappa$  to depend on  $h$ . We consider the extension of our results to this case in the next section.



We now are able to measure the complexity of computing an approximate solution to (2.1), with error not exceeding  $\epsilon$ , using a strong basic sequence  $\Phi$ . Indeed, (2.4) implies that a necessary and sufficient condition for  $\sigma(\varphi_p, h) = \epsilon$  is that

$$(2.5) \quad h = h(p, \alpha) := \kappa(p)^{-1/p} e^{-\alpha/p},$$

where

$$(2.6) \quad \alpha := \ln(\epsilon^{-1}).$$

(Note that since  $0 < \epsilon < 1$ , we have  $\alpha \in \mathbb{R}^{++}$ .) Thus, the number of steps needed is given by

$$(2.7) \quad n := h^{-1} = \kappa(p)^{1/p} e^{\alpha/p}.$$

(Note that  $n$  (as given by (2.7)) need not be an integer. But this poses no essential difficulty; see (e.g.) Traub and Woźniakowski [76].) Next, suppose that there exists an analytic function  $c: \mathbb{R}^+ \rightarrow \mathbb{R}^+$  such that  $c(p)$  is the cost per step associated with the method  $\varphi_p$ . Finally, we assume that the cost per step does not vary from step to step; for the classes of methods we consider, this means only that we assume that the cost of evaluating  $v$  (or its derivatives) does not depend on the point of evaluation. Thus the complexity  $C(p, \alpha)$  of solving (2.1) to within an error criterion  $\epsilon = e^{-\alpha}$  is simply given by

$$(2.8) \quad C(p, \alpha) = n c(p) = f(p) e^{\alpha/p},$$

where we define  $f: \mathbb{R}^+ \rightarrow \mathbb{R}^+$  by

$$(2.9) \quad f(p) := \kappa(p)^{1/p} c(p).$$

We now turn to the question of picking for each  $\alpha \in \mathbb{R}^{++}$  that order  $p$  giving minimal complexity. In the analysis to follow, we will drop the restriction that  $p$  must be an integer. However, we will recover optimality over the integers from optimality over the real numbers in Corollary 2.1. Without loss of generality, we assume that

$$(2.10) \quad p > 0 \text{ implies } f(p) > 0.$$

(If there were a  $p > 0$  with  $f(p) = 0$ , use of the method  $\varphi_p$  would yield a solution with zero complexity, i.e., "with no effort.") In addition, we assume that

$$(2.11) \quad \lim_{p \uparrow \infty} f(p) = +\infty .$$

By (2.9), this assumption maybe viewed as a simple consequence of two conditions, both of which are quite natural. The first is that  $\lim_{p \uparrow \infty} c(p) = +\infty$ ; the "better" a method is (i.e., the higher its order is), the more we should expect to pay for its use. The second condition is that if  $\lim_{p \uparrow \infty} \kappa(p) = 0$ , then there must exist a  $\beta \in I$  such that  $\kappa(p) \geq \beta^p$  for  $p$  sufficiently large. (For example, in the class of Taylor series methods, using the worst-case local error per unit step as the error measure, this second condition would follow from the assumption that any problem  $(x_0, v) \in \mathcal{D} \times \mathcal{V}$  must have an analytic solution.)

Thus in order to find a minimum for  $C(\cdot, \alpha)$ , we merely differentiate (2.8) with respect to  $p$ , finding

$$(2.12) \quad \partial_1 C(p, \alpha) = p^{-2} f(p) e^{\alpha/p} [G(p) - \alpha] ,$$

where  $G: \mathbb{R}^{++} \rightarrow \mathbb{R}$  is given by

$$(2.13) \quad G(p) := p^2 f'(p)/f(p) .$$

Thus a necessary condition that  $p$  be a minimum for  $C(\cdot, \alpha)$  is that  $\partial_1 C(p, \alpha) = 0$ , i.e.,

$$(2.14) \quad G(p) = \alpha .$$

Sufficient conditions for the existence and uniqueness of a  $p$  satisfying (2.14) and minimizing  $C(\cdot, \alpha)$  are given in

Theorem 2.1: Let  $f$  satisfy (2.10) and (2.11). Suppose also that

$$(2.15) \quad G'(p) > 0 \text{ whenever } G(p) > 0.$$

Then there is a function  $p^*: \mathbb{R}^{++} \rightarrow \mathbb{R}^{++}$  such that (2.14) holds if and only if  $p = p^*(\alpha)$ .

Moreover, for all  $p \in \mathbb{R}^{++}$ ,

$$(2.16) \quad C^*(\alpha) := C(p^*(\alpha), \alpha) \leq C(p, \alpha),$$

with equality holding if and only if  $p = p^*(\alpha)$ .

(Since  $p^*(\alpha)$  satisfies (2.16), we call  $p^*(\alpha)$  the optimal order,  $C^*(\alpha)$  the optimal complexity, and

$$(2.17) \quad h^*(\alpha) := h(p^*(\alpha), \alpha)$$

the optimal step-size.)

Proof of Theorem 2.1: If we write the Maclaurin series of  $f$  and substitute it into (2.13), it is easy to see that

$$(2.18) \quad \lim_{p \downarrow 0} G(p) = 0.$$

We now claim that

$$(2.19) \quad \lim_{p \uparrow \infty} G(p) = +\infty.$$

Indeed, since (2.11) holds there is a  $p_0 > 0$  such that  $f'(p_0) > 0$ , i.e.,  $G(p_0) > 0$ . Thus by (2.15),  $G$  is monotone increasing on  $[p_0, +\infty)$ , and hence either (2.19) holds or there exists a  $\gamma > 0$  such that  $\lim_{p \uparrow \infty} G(p) = \gamma$ . If the latter holds, then  $G$  is bounded, and we have

$$f'(t)/f(t) \leq \delta t^{-2} \quad (1 \leq t < +\infty)$$

for some  $\delta > 0$ ; integrating the above inequality over  $1 \leq t \leq p$  yields

$$f(p) \leq f(1) e^{\delta(1 - 1/p)},$$

so that  $\lim_{p \uparrow \infty} f(p) \leq f(1) e^{\delta}$ , contradicting (2.11). Thus (2.18) and (2.19) hold; together, they imply that for any  $\alpha > 0$ , there is a choice of  $p$  such that (2.14) holds.

Suppose that for some  $\alpha > 0$ , there were two numbers  $p_0 < p_1$  with  $G(p_0) = G(p_1) = \alpha$ . Then by Rolle's Theorem, there is a  $p_2$  between  $p_0$  and  $p_1$  with  $G'(p_2) = 0$ , contradicting (2.15). Thus for each  $\alpha > 0$ , there is a unique choice of  $p$  such that (2.14) holds; we denote this choice by  $p^*(\alpha)$ .

To prove (2.16), differentiate (2.12) with respect to  $p$  to find

$$(2.20) \quad \partial_1^2 C(p, \alpha) = p^{-2} f(p) e^{\alpha/p} G'(p) + [G(p) - \alpha] (\partial/\partial p) [p^{-2} f(p) e^{\alpha/p}] .$$

But upon substituting  $p = p^*(\alpha)$ , the second term in (2.20) vanishes and the first term is positive; so we have

$$\partial_1^2 C(p^*(\alpha), \alpha) > 0 .$$

Thus  $p^*(\alpha)$  gives a local minimum for  $C(\cdot, \alpha)$ , which has only one critical point (since (2.14) has a unique solution) and (2.16) follows. ■

Note that we have not said that  $p^*(\alpha)$  is an integer; in fact, this need not be true in general. Since the basic sequence  $\Phi$  is indexed by  $\mathbb{Z}^{++}$ , we have not yet solved the problem of choosing from among all  $(\varphi_p, h)$  such that (2.3) holds, that pair yielding minimal complexity. This problem is solved by

Corollary 2.1: For any  $\alpha > 0$ , define  $p^{**}(\alpha) \in \mathbb{Z}^{++}$  to be that element of the set  $\{[p^*(\alpha)], \lceil p^*(\alpha) \rceil\}$  which gives the smaller value of  $C(\cdot, \alpha)$ . Then

$$C(p^{**}(\alpha), \alpha) \leq C(p, \alpha) \quad \text{for } p \in \mathbb{Z}^{++} ,$$

with equality if and only if  $p = p^{**}(\alpha)$ .

Proof: Clearly we need only consider the case where  $p^*(\alpha)$  is not an integer. Suppose there exists  $p_0 \in \mathbb{Z}^{++}$ , not equal to  $p^{**}(\alpha)$ , with  $C(p_0, \alpha) \leq C(p^{**}(\alpha), \alpha)$ . Without loss of generality, assume  $p_0 < [p^*(\alpha)]$ . Then  $C(p_0, \alpha) \leq C([p^*(\alpha)], \alpha) \geq C(p^*(\alpha), \alpha)$ , which implies that there is a  $p_1 \in (p_0, p^*(\alpha))$  such that  $\partial_1 C(p_1, \alpha) = 0$ . Hence,  $G(p_1) = \alpha$ , but  $p_1 \neq p^*(\alpha)$ . This contradicts Theorem 2.1. ■

It may be readily verified that the hypotheses of Theorem 2.1 are satisfied for many classes of functions  $f$ . Some of these are

logarithmic:  $f(p) = \ln(p + e)$ ,

monomial:  $f(p) = p^m$  ( $m \in \mathbb{R}^{++}$ ),

exponential:  $f(p) = \beta^p$  ( $\beta > 1$ ),

super-exponential:  $f(p) = p^p$ , and

hyper-exponential:  $f(p) = \beta^{p^p}$ .

(We write " $\ln(p + e)$ ", where  $e$  is the base of the natural logarithms, rather than " $\ln p$ " as a technical convenience. However, an expression of the form " $\ln(p + \gamma)$ " with  $\gamma > 0$  is necessary to guarantee that  $f(1) > 0$ .) Furthermore, we find that if  $f$  has the monomial-logarithmic form

$$f(p) = p^a (\ln(p+e))^b \quad (a, b \in \mathbb{R}^{++}),$$

then the hypotheses of Theorem 2.1 hold. This may be verified either directly, or by using the following Lemma, along with the fact that the hypotheses hold for  $f(p) = p$  and  $f(p) = \ln(p + e)$ .

Lemma 2.1: Let  $f$  have the form

$$f(p) = a \prod_{i=1}^m (f_i(p))^{r_i},$$

where  $a \in \mathbb{R}^{++}$ , and for each  $i$  ( $1 \leq i \leq m$ ),  $f_i$  satisfies the hypotheses of Theorem 2.1 and  $r_i \in \mathbb{R}^{++}$ . Then  $f$  satisfies the hypotheses of Theorem 2.1.

Proof: It is clear that if each  $f_i$  satisfies (2.10) and (2.11), then so does  $f$ . If each  $f_i$  yields (via (2.13)) a  $G_i$  satisfying (2.15), then  $f$  yields a  $G$  in the form

$$G(p) = \sum_{i=1}^m r_i G_i(p),$$

and so it is clear that  $G$  satisfies (2.15). ■

For the important methods of practical interest, we will only be interested in

monomial and monomial-logarithmic growth; see Werschulz [76a], [76b]. We include the other examples of functions that satisfy the hypotheses of Theorem 2.1 to illustrate the wide variety of functions that qualify.

So we have seen that under the hypotheses of Theorem 2.1, there is a unique choice of order and step size minimizing the total complexity for any error criterion. What happens to these choices as  $\alpha$  changes?

Theorem 2.2: Let  $f$  satisfy the hypotheses of Theorem 2.1. Then

- (1.)  $p^*(\alpha)$  and  $C^*(\alpha)$  increase monotonically with  $\alpha$ .
- (2.)  $\lim_{\alpha \uparrow \infty} p^*(\alpha) = \lim_{\alpha \uparrow \infty} C^*(\alpha) = +\infty$ .
- (3.) If there exists  $M > 0$  such that  $\kappa(p)^{1/p} \leq M$  for all  $p$ , then  $\liminf_{\alpha \uparrow \infty} h^*(\alpha) > 0$  if  $\alpha/p^*(\alpha)$  is bounded as  $\alpha \uparrow \infty$ .

Proof: To prove (1.), note that  $p^*$  is the functional inverse of  $G$ . Thus  $p^{*'}(\alpha) = G'(p^*(\alpha))^{-1} > 0$ , so that  $p^*(\alpha)$  increases with  $\alpha$ . Now use the chain rule:

$$C^{*'}(\alpha) = \partial_1 C(p^*(\alpha), \alpha) p^{*'}(\alpha) + \partial_2 C(p^*(\alpha), \alpha).$$

But the first term on the right-hand side vanishes by the definition of  $p^*(\alpha)$ . So

$$C^{*'}(\alpha) = \partial_2 C(p^*(\alpha), \alpha) = (p^*(\alpha))^{-1} f(p^*(\alpha)) e^{\alpha/p^*(\alpha)} > 0$$

and  $C^*(\alpha)$  increases with  $\alpha$ .

Suppose that  $\lim_{p \uparrow \infty} p^*(\alpha) \neq +\infty$ . Since  $p^*(\alpha)$  increases monotonically with  $\alpha$ , there is an  $L > 0$  such that  $\lim_{\alpha \uparrow \infty} p^*(\alpha) = L$ . So (2.14) implies that

$$G(L) = \lim_{\alpha \uparrow \infty} G(p^*(\alpha)) = \lim_{\alpha \uparrow \infty} \alpha = +\infty,$$

contradicting the continuity of  $G$ . This proves the first part of (2.). Now for any  $\alpha > 0$ , we have

$$C^*(\alpha) = f(p^*(\alpha)) e^{\alpha/p^*(\alpha)} > f(p^*(\alpha)).$$

Let  $\alpha \uparrow \infty$ ; then (2.11) and the first part of (2.) imply that the second part of (2.) holds.

To prove (3.), let such an  $M > 0$  exist, so that  $[h^*(\alpha)]^{-1} \leq M e^{\alpha/p^*(\alpha)}$ . Then we

see that  $\liminf_{\alpha \uparrow \infty} h^*(\alpha) > 0$  if  $[h^*(\alpha)]^{-1}$  is bounded as  $\alpha \uparrow \infty$ , which itself is true if  $\alpha/p^*(\alpha)$  is bounded as  $\alpha \uparrow \infty$ . ■

Therefore under a very general set of conditions, we see that the more accuracy we want in our computed solution, the greater its complexity becomes. Of course, this is just what we would expect. What is somewhat surprising is that the minimal complexity is obtained by letting the order  $p$  increase as the error  $\epsilon$  decreases, with  $p$  increasing without bound as  $\epsilon$  tends to zero. Moreover, the last part of the theorem says that not only should the order be increased when trying to obtain a more accurate solution, but that it may actually turn out that the step-size should not be allowed to tend to zero. In addition, it is clear that the proof of the result concerning the limiting behavior is valid, provided that we only assume that  $f$  is continuously differentiable on the positive real axis and that  $p^*(\alpha)$  tends to a (possibly infinite) limiting value as  $\alpha \uparrow \infty$ .

We now determine whether we are saving a great deal by using the optimal-order method. This may be thought of in several ways; we will consider how sharply the cost curve turns at the optimum, the cost-difference between using a method of fixed order and a method of optimal order, and the cost-ratio of a fixed-order method to an optimal-order method. We will show that under certain reasonable conditions, all of these measures tend to infinity with  $\alpha$ .

How sharply the cost curve turns at the maximum is measured by  $\partial_1^2 C(p^*(\alpha), \alpha)$ . If we consider five of the growth models mentioned above (e.g., monomial, monomial-logarithmic, exponential, hyper-exponential, and super-exponential), we find that  $\partial_1^2 C(p^*(\alpha), \alpha)$  is monotone increasing for  $\alpha$  sufficiently large, and tends to infinity with  $\alpha$ , with but one exception; in the case of "linear growth" ( $f(p) = p$ ), we find that

$\partial_1^2 C(p^*(\alpha), \alpha) = e$ . However, in the classes of algorithms we study, the case  $f(p) = p$  does not arise, provided that we include "combinatory cost" (defined as in Kung and Traub [74]) in our complexity measure. Thus in general, we find that the "pointedness" of the cost curve near the minimum increases without bound as  $\alpha \uparrow \infty$ .

Next, we will show that for any  $f$  satisfying the hypotheses of Theorem 2.1, the difference in complexity between using a method of fixed order and a method of optimal order tends to infinity with  $\alpha$ .

Proposition 2.2: For any fixed  $p_0 \in \mathbb{R}^{++}$  such that  $G'(p_0) \geq 0$ ,

$$\lim_{\alpha \uparrow \infty} [C(p_0, \alpha) - C^*(\alpha)] = +\infty.$$

Proof: Pick  $\alpha$  so large that  $p^*(\alpha) > p_0$ , and let  $p_0 < p < p^*(\alpha)$ . If we write out the partial derivative in the last term of (2.20), we find that

$$\partial_1^2 C(p, \alpha) = p^{-2} f(p) e^{\alpha/p} G'(p) + p^{-4} [\alpha - G(p)] [(\alpha + 2p) - G(p)] f(p).$$

Since  $p_0 < p < p^*(\alpha)$ , we have  $G(p) < \alpha$ ; it then follows that  $\partial_1^2 C(p, \alpha)$  is positive and bounded away from zero as  $\alpha$  tends to infinity. Since

$$C(p_0, \alpha) - C^*(\alpha) = \partial_1^2 C(p, \alpha) [p_0 - p^*(\alpha)]^2 / 2$$

for some  $p$  between  $p_0$  and  $p^*(\alpha)$ , the result follows. ■

As for the cost-ratio, a simple calculation shows that

$$\lim_{\alpha \uparrow \infty} C(p_0, \alpha) / C^*(\alpha) = +\infty$$

in all of the examples given above. Thus there are a number of ways in which we incur a large additional cost by not using the optimal order.

One may wonder whether the result that optimal order increases and tends to infinity with  $\alpha$  is "reasonable." One way of determining this is to examine actual numerical tests; we cite Hull et al. [72] as a well-known example. Since we are only dealing with methods of fixed order, our theory does not attempt to handle methods



such as Bulirsch-Stoer, Krogh, or Gear. However, let us look at the results of Hull et al. for the Runge-Kutta methods. Even though there are only three methods (of orders four, six, and eight) and three error criteria ( $\epsilon = 10^{-3}$ ,  $10^{-6}$ , and  $10^{-9}$ ), Table 1 in Hull et al. [72] indicates that the optimal order does increase as  $\epsilon$  decreases. (We give more extensive numerical data in Section 5.)

Finally, we note that the restriction that the grid be equidistant may be weakened somewhat, provided that we use a local error measure. Indeed, let  $I$  be partitioned as  $I = I_1 \cup \dots \cup I_L$ , and now assume that we use a grid that is equidistant on each subinterval  $I_1, \dots, I_L$ . Then the total complexity is given by the sum of the complexities of all subintervals

$$C(p_1, \dots, p_L, \alpha) := \sum_{i=1}^L C_i(p_i, \alpha),$$

where we set

$$C_i(p, \alpha) := f_i(p, \alpha) e^{\alpha/p}, \quad f_i(p) := \kappa_i(p)^{1/p} c(p);$$

here  $\kappa_i(p)$  is the error constant of  $\varphi_p$  on  $I_i$ . Since we use a local error measure, we find that  $C(p_1, \dots, p_L, \alpha)$  is minimized by choosing each  $p_i$  to minimize  $C_i(\cdot, \alpha)$ . Thus the earlier results apply; in particular, if we define  $p_i^*(\alpha)$  to be the optimal order on  $I_i$ , we find that if  $f_i$  satisfies (2.10), (2.11), and (2.15), then  $p_i^*(\alpha)$  increases and tends to infinity with  $\alpha$ .

### 3. Optimality Within a Basic Sequence

There are two difficulties with the approach taken in Section 2. The first has already been mentioned--we generally expect the error coefficient to depend on the step-size. The second is based on the fact that there are a large number of  $p^{\text{th}}$ -order methods of a given type, and we wish to use the best method possible. In theory, this would involve finding a  $p^{\text{th}}$ -order method with minimal cost per step. In practice, this is not often possible; there is a gap between the minimal cost theoretically possible and the cost of the best method known. So we now consider the extension of the results in Section 2 to a more general setting, which will take these two difficulties into account.

We first refine our notion of order. Let  $\sigma: \Phi \times I \rightarrow \mathbb{R}^+$  be an error measure, where  $\Phi = \{\varphi_p: p \in \mathbb{Z}^{++}\}$  is a class of one-step methods, and suppose that a function  $\kappa: \mathbb{R}^+ \times I \rightarrow \mathbb{R}^+$  and analytic functions  $\kappa_L, \kappa_U: \mathbb{R}^+ \rightarrow \mathbb{R}^+$  exist such that  $\lim_{p \rightarrow 0} \kappa_L(p)^{1/p}$  and  $\lim_{p \rightarrow 0} \kappa_U(p)^{1/p}$  exist and are nonzero and

$$(3.1) \quad \sigma(\varphi_p, h) = \kappa(p, h) h^p \quad \text{for } h \in I \text{ and } p \in \mathbb{Z}^{++},$$

where

$$(3.2) \quad 0 < \kappa_L(p) \leq \kappa(p, h) \leq \kappa_U(p) < +\infty \quad \text{for } h \in I.$$

Then  $\varphi_p$  is said to have order  $p$  with respect to  $\sigma$ , and  $\Phi$  is said to be a basic sequence (as in Traub [64]);  $\kappa(p, h)$  is said to be the error coefficient of  $\varphi_p$ . (Here we introduce the convention of attaching the subscripts "L" and "U" to quantities that refer to lower and upper bounds on complexity, respectively.)

This definition of order is similar to that in Cooper [69] and Cooper and Verner [72], except that we include a lower bound  $\kappa_L(p)$  on  $\kappa(p, h)$ ; this lower bound is

necessary and sufficient to guarantee that the order of a method is well-defined. Note that this definition makes sense for all values of  $h \in \mathbb{I}$ ; thus, it is non-asymptotic in that we do not require  $h \downarrow 0$  in order for it to make sense. Clearly, a strong basic sequence is a basic sequence; hence, the definition of order is an extension of the definition of strong order given in Section 2. Finally, note that the order depends on the choice of the error measure  $\sigma$ ; for instance, the order with respect to the local error per step exceeds that with respect to the local error per unit step by one.

We next discuss the notion of cost per step. As pointed out above, we will generally have only bounds on the cost  $c(p)$  required per step of a given  $p^{\text{th}}$ -order method:

$$(3.3) \quad c_L(p) \leq c(p) \leq c_U(p).$$

That is,  $c_L(p)$  is a lower bound on the minimum possible cost per step, usually derived via theoretical considerations, and  $c_U(p)$  is an upper bound on the minimum possible cost per step, which is derived by exhibiting an algorithm for computing  $\varphi_p$ . (In what follows, we shall assume that  $c_L, c_U : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  are analytic functions.)

We now wish to give bounds on  $C(p, \alpha)$ , the complexity of finding an approximate solution of (2.1) using the method  $\varphi_p$ , such that  $\sigma(\varphi_p, h) \leq e^{-\alpha}$ . Suppose that (2.3) holds. Then by (3.1) and (3.2), we must have

$$(3.4) \quad \kappa_L(p)h^p \leq e^{-\alpha}, \quad \text{i.e.,} \quad h \leq h_L(p, \alpha) := \kappa_L(p)^{-1/p} e^{-\alpha/p}.$$

Hence, the number of steps  $n = h^{-1}$  must satisfy

$$(3.5) \quad n \geq \kappa_L(p)^{1/p} e^{\alpha/p}.$$

Defining (as in Section 2)

$$(3.6) \quad C(p, \alpha) := n c(p)$$

(i.e., total complexity equals number of steps required multiplied by cost required per step), (3.3) and (3.5) imply that

$$(3.7) \quad C(p, \alpha) \geq C_L(p, \alpha) := f_L(p) e^{\alpha/p},$$

where

$$(3.8) \quad f_L(p) := \kappa_L(p)^{1/p} c_L(p).$$

That is, regardless of the algorithm used to compute  $\varphi_p$ , the total complexity of finding an approximate solution of (2.1) must exceed  $C_L(p, \alpha)$ .

On the other hand, we find that in order to use  $\varphi_p$  to find such an approximate solution, it suffices (by (3.1) and (3.2)) to take

$$(3.9) \quad \kappa_U(p) h^p = e^{-\alpha}, \text{ i.e., } h = h_U(p, \alpha) := \kappa_U(p)^{-1/p} e^{-\alpha/p}.$$

so that we need only take  $n$  steps, where

$$(3.10) \quad n = \kappa_U(p)^{1/p} e^{\alpha/p}.$$

(As in Section 2, the value of  $n$  given by (3.10) need not be an integer; again, this is handled as in Traub and Woźniakowski [76].) Thus (3.3), (3.6), and (3.10) imply that

$$(3.11) \quad C(p, \alpha) \leq C_U(p, \alpha) := f_U(p) e^{\alpha/p},$$

where

$$(3.12) \quad f_U(p) := \kappa_U(p)^{1/p} c_U(p).$$

That is, there exists an algorithm for computing  $\varphi_p$  such that the total complexity of finding an approximate solution of (2.1) equals  $C_U(p, \alpha)$ . We summarize the above results in

**Theorem 3.1:** Let  $C(p, \alpha)$  be the complexity of finding an approximate solution of (2.1), using the method  $\varphi_p$ , with  $\sigma(\varphi_p, h) \leq e^{-\alpha}$ . Then

$$(3.13) \quad C_L(p, \alpha) \leq C(p, \alpha) \leq C_U(p, \alpha),$$

where  $C_L$  and  $C_U$  are given by (3.7) and (3.11). Moreover, if  $h = h(p, \alpha)$  is the maximal step-size for the method  $\varphi_p$  such that  $\sigma(\varphi_p, h) \leq e^{-\alpha}$ , then

$$(3.14) \quad h_U(p, \alpha) \leq h(p, \alpha) \leq h_L(p, \alpha). \quad \blacksquare$$

Next, we consider the problem of optimality. Define the optimal complexity by

$$(3.15) \quad C^*(\alpha) := \inf \{C(p, \alpha) : \varphi_p \in \Phi\}.$$

We are interested in bounds for  $C^*(\alpha)$ . These are derived in

Lemma 3.1: Let  $f_L$  and  $f_U$  satisfy (2.10) and (2.11), and suppose that  $f_L$  and  $f_U$  respectively yield (via (2.15))  $G_L$  and  $G_U$  satisfying (2.15). Then  $G_L$  and  $G_U$  have respective inverse functions  $p_L^*, p_U^*: \mathbb{R}^{++} \rightarrow \mathbb{R}^{++}$  such that for all  $p \in \mathbb{R}^{++}$ ,

$$(3.16) \quad C_L^*(\alpha) := C_L(p_L^*(\alpha), \alpha) \leq C_L(p, \alpha)$$

and

$$(3.17) \quad C_U^*(\alpha) := C_U(p_U^*(\alpha), \alpha) \leq C_U(p, \alpha),$$

with equality in (3.16) (respectively, (3.17)) if and only if  $p = p_L^*(\alpha)$  (respectively,  $p = p_U^*(\alpha)$ ).

Proof: This is an immediate corollary of Theorem 2.1. ■

We call  $p_L^*(\alpha)$  (respectively,  $p_U^*(\alpha)$ ) the lower (upper) optimal order,  $C_L^*(\alpha)$  (respectively,  $C_U^*(\alpha)$ ) the lower (upper) optimal complexity, and

$$(3.18) \quad h_L^*(\alpha) := h_L(p_L^*(\alpha), \alpha) \quad (\text{respectively, } h_U^*(\alpha) := h_U(p_U^*(\alpha), \alpha))$$

the lower (upper) optimal step-size. Combining (3.13), (3.15), and Lemma 3.1, we have

Theorem 3.2: Let  $f_L$  and  $f_U$  be as in Theorem 3.1. Then

$$C_L^*(\alpha) \leq C^*(\alpha) \leq C_U^*(\alpha). \quad \blacksquare$$

Note that if we define  $p^*(\alpha)$  by

$$C(p^*(\alpha), \alpha) = C^*(\alpha),$$

we can make no statement relating  $p^*(\alpha)$ ,  $p_L^*(\alpha)$ , and  $p_U^*(\alpha)$ . This is because we only have bounds for  $C(p, \alpha)$ ; we do not know  $C(p, \alpha)$  itself. In fact, it is important to realize what  $p_L^*(\alpha)$  and  $p_U^*(\alpha)$  tell us. First, consider  $p_U^*(\alpha)$ . We can achieve a complexity of  $C_U^*(\alpha)$  by using a step-size of  $h_U^*(\alpha)$ , along with the method of order  $p_U^*(\alpha)$ . This will give optimal complexity within the sequence of algorithms for computing  $\Phi$ , with cost

per step of  $\varphi_p$  given by  $c_U(p)$ . Next, consider  $p_L^*(\alpha)$ . It is of perhaps theoretical rather than computational interest, in that we cannot compute with it. What does interest us is  $C_L^*(\alpha)$ , since it limits the theoretical improvement in  $C_U^*(\alpha)$ . Thus, we are interested in  $p_L^*(\alpha)$  solely as a means of computing  $C_L^*(\alpha)$ .

We now consider behavior of these quantities as  $\alpha$  increases and tends to infinity.

Theorem 3.3: Let  $f_L$  and  $f_U$  be as in Theorem 3.1. Then

- (1.)  $p_L^*(\alpha)$ ,  $p_U^*(\alpha)$ ,  $C_L^*(\alpha)$ , and  $C_U^*(\alpha)$  increase monotonically and tend to infinity with  $\alpha$ .
- (2.) If there exists an  $M_U > 0$  such that  $\kappa_U(p)^{1/p} \leq M_U$  for all  $p$ , then  $\liminf_{\alpha \uparrow \infty} h_U^*(\alpha) > 0$  if  $\alpha/p_U^*(\alpha)$  is bounded as  $\alpha \uparrow \infty$ .
- (3.) If there exists an  $M_L > 0$  such that  $\kappa_L(p)^{1/p} \geq M_L$  for all  $p$ , then  $\liminf_{\alpha \uparrow \infty} h_L^*(\alpha) > 0$  only if  $\alpha/p_L^*(\alpha)$  is bounded as  $\alpha \uparrow \infty$ .

Proof: To prove (1.), it suffices to apply (1.) and (2.) of Theorem 2.2 to  $p_L^*$  and  $C_L^*$ , and to  $p_U^*$  and  $C_U^*$ . The proof of (2.) and (3.) is similar to the proof of (3.) in Theorem 2.2. ■

Note that (1.) in Theorem 3.3 does not state how  $p^*(\alpha)$  varies with  $\alpha$ ; as we have pointed out above, no statement about  $p^*(\alpha)$  may be obtained from the information available. However, it is easy to see that  $C^*(\alpha)$  increases monotonically with  $\alpha$  and that  $\lim_{\alpha \uparrow \infty} C^*(\alpha) = +\infty$ .

Thus, we have extended the optimality theory of Section 2 to a more realistic situation. In Werschulz [76a], [76b], the techniques of this section are applied to some important basic sequences of one-step methods; we will see that the conclusions of Lemma 3.1 and Theorems 3.2 and 3.3 hold for these basic sequences.

## 4. Normality and Order-Convergence

Let  $\Phi$  be a basic sequence with respect to the error measure  $\sigma$ ; we say that  $\Phi$  is order-convergent if there exists an  $h_0 > 0$  such that

$$(4.1) \quad \lim_{p \uparrow \infty} \kappa_U(p) h^p = 0 \quad \text{for } h \leq h_0 .$$

Clearly, the order convergence of  $\Phi$  implies that  $\lim_{p \uparrow \infty} \sigma(\varphi_p, h) = 0$  for  $h \leq h_0$ . We use the term "order-convergence" rather than "convergence," since the latter term appears extensively in the literature (e.g., Henrici [62]) and is always used to mean a "step-size convergence," i.e.,  $\lim_{h \downarrow 0} \sigma(\varphi, h) = 0$  for a fixed method  $\varphi$ .

It is intuitively plausible that as the order of an approximation increases, the approximation should improve, especially when one is trying to approximate a very smooth function. Unfortunately, Gear [71] points out that an increase in order need not always decrease the error. This situation appears in other situations in numerical mathematics; for instance, the family of Newton-Cotes quadrature formulae is not order-convergent. But suppose there exists a step-size  $h_0 > 0$  for which the upper-bound error is exponentially bounded for  $p$  sufficiently large; that is, there exists  $A > 0$  and  $p_0 \in \mathbb{Z}^{++}$  such that

$$(4.2) \quad \kappa_U(p) h_0^p \leq A^p \quad \text{for } p > p_0 .$$

If we define

$$M_U := \max \{ \max_{1 \leq p \leq p_0} \{ \kappa_U(p)^{1/p} \}, A h_0^{-1} \},$$

we then have

$$(4.3) \quad \sigma(\varphi_p, h) \leq (M_U h)^p \quad \text{for } h \leq h_0, p \in \mathbb{Z}^{++} .$$

Note that the bound in (4.3) is similar to that given by Cauchy's Integral Theorem (Ahlfors [66], pg. 122) on the normalized derivatives of an analytic function. In fact,

for several classes of methods, the bound (4.3) holds whenever the solution of (2.1) is analytic.

We also formalize a weakened version of (4.3), which will be important in our study of one-step methods. Let  $\Phi$  be a basic sequence, and suppose that for each  $(x_0, v) \in \mathcal{D} \times \mathcal{V}$ , there is a sequence  $\{h_p: p \in \mathbb{Z}^{++}\} \subset I$  and a positive constant  $M_U$  such that

$$(4.4) \quad \sigma(\varphi_p, h) \leq (M_U h)^p \quad \text{if } h \leq h_p;$$

then  $\Phi$  is said to be normal. Note that (4.3) implies (4.4), while (4.4) implies (4.3) only when the sequence  $\{h_p\}$  has non-vanishing support:

$$(4.5) \quad h_\Phi := \liminf_{p \uparrow \infty} h_p > 0.$$

If  $h_\Phi = 0$ , normality gives an exponential upper bound on the sequence of principal error functions (Section 3.3-5 of Henrici [62]), which are an asymptotic measure of the error as  $h \downarrow 0$ .

There is a simple relation between normality and order-convergence.

Proposition 4.1:  $\Phi$  is order-convergent if and only if  $\Phi$  is normal with nonvanishing support.

Proof: If (4.1) holds, then (in particular) we have  $\lim_{p \uparrow \infty} \kappa_U(p) h_0^p = 0$ , so that  $\kappa_U(p) h_0^p \leq 1$  for  $p$  sufficiently large; i.e., (4.2) holds with  $A = 1$ . Then (as in the discussion above) (4.3) holds, implying normality with finite support.

Conversely, if (4.4) holds with finite support, we pick a positive  $h_0$  which is less than

$$\eta := \min \{M_U^{-1}, \inf \{h_p: p \in \mathbb{Z}^{++}\}\}.$$

(Note that  $\eta > 0$  by (4.5).) Let  $h \leq h_0$  be given, so that for some  $\delta$  with  $0 < \delta < 1$ , we have  $h = (1 - \delta)\eta$ ; if we define  $\kappa_U$  by

$$\kappa_U(p) := M_U^p,$$



we find (since  $h < h_p$ ) that

$$\sigma(\varphi_p, h) \leq (M_U h)^P .$$

Thus

$$\kappa_U(p) h^P = (M_U h)^P = (M_U (1 - \delta) \eta)^P \leq (1 - \delta)^P$$

(the last step since  $\eta \leq M_U^{-1}$ ), so that (4.1) holds. ■

We are now interested in normality and order-convergence for a specific error measure  $\sigma$ ; we will be interested in  $\sigma_{LU}$ ,  $\sigma_L$ , and  $\sigma_G$ , which are (respectively) defined to be the maximum local error per unit step, local error per step, and global error per step over the grid. It is easy to see that a normal (order-convergent) sequence  $\Phi = \{\varphi_p: p \in \mathbb{Z}^{++}\}$  with respect to  $\sigma_L$  naturally yields a normal (order-convergent) sequence  $\Psi = \{\psi_p: p \in \mathbb{Z}^{++}\}$  with respect to  $\sigma_{LU}$  by setting  $\psi_p := \varphi_{p+1}$  for  $p \in \mathbb{Z}^{++}$ . We now look at the relationships between  $\sigma_{LU}$  and  $\sigma_G$ .

Proposition 4.2: Let  $v$  have Lipschitz constant  $K$  on  $\mathbb{R}^N$ , and let  $\Phi$  be normal (respectively, order-convergent) with respect to  $\sigma_{LU}$ , with  $M_U$  in (4.4) independent of  $x_0 \in \text{domain}(v)$ . Then  $\Phi$  is normal (respectively, order-convergent) with respect to  $\sigma_G$ .

Proof: Let  $\rho$  be the exact relative increment function of (2.1) (as defined in Henrici [62]), so that

$$x(t_{i+1}) = x(t_i) + h \rho(x(t_i), h) .$$

Subtract (2.2) (with  $\varphi$  replaced by  $\varphi_p$ ) from the above to get

$$e_{i+1} = e_i + h [\rho(x(t_i), h) - \varphi_p(x_i, h)] ,$$

where  $e_i := x(t_i) - x_i$  for  $0 \leq i \leq n$ . Thus

$$\begin{aligned} \|e_{i+1}\| &\leq \|e_i\| + h \|\rho(x(t_i), h) - \rho(x_i, h)\| + h \|\rho(x_i, h) - \varphi_p(x_i, h)\| \\ &\leq (1 + hK) \|e_i\| + M_U^P h^{P+1} \quad \text{if } h \leq h_p ; \end{aligned}$$

this last step follows from the Lipschitz condition and the "uniform" normality with respect to  $\sigma_{LU}$ . By Lemma 1.2 of Henrici [62] and the condition  $e_0 = 0$ , we have

$$\begin{aligned}
\|e_i\| &\leq K^{-1} [(1 + hK)^i - 1] (M_U h)^p \\
&\leq K^{-1} [(1 + hK)^n - 1] (M_U h)^p \\
&\leq K^{-1} (e^K - 1) (M_U h)^p
\end{aligned}$$

for all  $i$ ; this gives

$$\sigma_G(\varphi_p, h) \leq K^{-1} (e^K - 1) (M_U h)^p \leq (Mh)^p \quad \text{if } h \leq h_p,$$

for a suitably-defined  $M > 0$ . This proves the normality part; the remainder of the result follows from Proposition 4.1. ■

If it is undesirable to use the "uniform normality" (i.e., the condition that  $M_U$  be independent of  $x_0 \in \text{domain}(v)$  in (4.4)), we may use the following result.

Proposition 4.3: Let  $v$  be Lipschitz continuous, let  $\Phi$  be normal (respectively, order-convergent) with respect to  $\sigma_{LU}$ , and suppose that there exists a  $\lambda > 0$  such that for all  $\varphi_p \in \Phi$  and all  $x, y \in \mathbb{R}^N$ ,

$$\|\varphi_p(x) - \varphi_p(y)\| \leq \lambda p \|x - y\|.$$

Then  $\Phi$  is normal (respectively, order-convergent) with respect to  $\sigma_G$ .

Proof: Immediate from Theorem 3.3 of Henrici [62]. ■

Thus normality for  $\sigma_G$  follows from normality for  $\sigma_{LU}$ , a Lipschitz condition on  $v$  and the elements of  $\Phi$ , and a linear upper bound on the Lipschitz constants for the elements of  $\Phi$ .

We now discuss the problem of finding uniform lower bounds on the error which are similar to the uniform upper bounds which normality provides. This will amount to a restriction of the admissible problem class  $\mathcal{D} \times \mathcal{V}$  so as to guarantee that the problems are "sufficiently difficult." However, this restriction may be abandoned if we are interested only in upper bounds. We shall assume throughout the rest of this section that there is an  $M_L > 0$  (which will generally depend on  $\Phi$ ,  $\sigma$ , and the problem  $(x_0, v)$ ) such that

$$(4.6) \quad \sigma(\varphi_p, h) \geq (M_L h)^p \quad \text{for } h \in I.$$

Note that (4.6) will hold for any situation in which there is no order-convergence, or in which the order-convergence (if any) is no faster than an exponential decay; moreover, in the methods we consider in Sections 5 and 6, (4.6) is a consequence of the assumption that all derivatives assume the (sharp) worst-case upper bound provided by Cauchy's estimate. It is clear that if (4.6) holds for  $\sigma_L$ , it holds for  $\sigma_{LU}$ ; if (4.6) holds for  $\sigma_{LU}$  and if the gradient matrix  $\nabla \rho$  has only non-negative entries (with at least one positive entry), then (4.6) holds for  $\sigma_G$ .

It is possible to present a simplified version of the expressions derived in Section 3, under the assumption that  $\Phi$  is order-convergent. We first look at the complexity of a single method within an order-convergent basic sequence.

Theorem 4.1: Let  $\Phi$  be order-convergent with respect to  $\sigma$ . Then

$$C_L(p, \alpha) \leq C(p, \alpha) \leq C_U(p, \alpha),$$

where

$$C_L(p, \alpha) := M_L c_L(p) e^{\alpha/p} \quad \text{and} \quad C_U(p, \alpha) := M_U c_U(p) e^{\alpha/p}.$$

Proof: This is an immediate corollary of Theorem 3.1 and the definition of order-convergence. ■

We may now do the optimality theory of Section 3, finding that

$$(4.7) \quad G_L(p) = p^2 c_L'(p)/c_L(p) \quad \text{and} \quad G_U(p) = p^2 c_U'(p)/c_U(p).$$

Note that the assumptions (2.10) and (2.11) now state that  $c_L(p)$  and  $c_U(p)$  must be positive for  $p > 0$  and tend to infinity with  $p$ , which is a natural way to expect the cost per step to behave. The results stated in Theorems 3.2 and 3.3 hold as before. Moreover, it should be noted that the  $M_U$  and  $M_L$  needed in (2.) and (3.) in the statement of Theorem 3.3 are precisely the  $M_U$  and  $M_L$  in (4.4) and (4.6). Thus

$\liminf_{\alpha \uparrow \infty} h_U^*(\alpha) > 0$  if  $\alpha/p_U^*(\alpha)$  is bounded as  $\alpha \uparrow \infty$ , and  $\alpha/p_L^*(\alpha)$  is bounded as  $\alpha \uparrow \infty$  if  $\liminf_{\alpha \uparrow \infty} h_L^*(\alpha) > 0$ .

Thus, the order-convergence of a basic sequence is useful in simplifying the analysis of its complexity. Of the three basic sequences studied in Werschulz [76a], [76b], two are known to be order-convergent. The proof of the order-convergence of the class of Taylor series methods is a simple consequence of the Cauchy estimate; that of the order-convergence of the (non-optimally ordered) nonlinear Brent-Runge-Kutta methods (based on the iterative methods defined in pp. 4-7 of Brent [74]) involves using some classical results on orthogonal polynomials to sharpen the proofs in Brent [74]. We note that it is not known whether the optimally-ordered nonlinear Brent-Runge-Kutta methods (based on the iterative methods defined in pp. 10-13 of Brent [74]) are order-convergent; it does appear likely that they are normal with vanishing support. However, we do not pursue this class of methods in Werschulz [76b], because of their high combinatory cost.

It is not known whether the linear Runge-Kutta methods found in Cooper [69] and in Cooper and Verner [72] are order-convergent; the best result known is the  $(M_U \log(p+\epsilon))^P$  result given in Werschulz [76a], which involves strengthening the original proof with other estimates from the theory of orthogonal polynomials. But it should be pointed out that there does exist a class of order-convergent linear Runge-Kutta methods; this is the sequence given by using the weights and abscissae for Gauss quadrature in the methods defined on page 144 of Stetter [73]. The problem with this class of methods is that each step of  $\varphi_p$  requires  $2^{-P(p+1)!}$  function evaluations; the prohibitive cost per step outweighs by far any advantage to be gained from the order-convergence. Thus, the question of whether there exist any order

convergent linear Runge-Kutta methods which are more efficient (i.e., have smaller cost per step) remains open.

## 5. Numerical Results

In the previous sections, we showed that the optimal order increases as the error criterion  $\epsilon$  decreases, tending to infinity as  $\epsilon$  tends to zero. Here, we consider actual numerical results of optimal order and minimal cost for various test problems and classes of methods; these results show that the optimal order does indeed exhibit the behavior indicated. (The optimal order for a given error criterion was determined by finding, for each method implemented, the coarsest mesh that allowed the error criterion to be satisfied; the resulting complexities were then compared to determine the optimal order.) The error measure used was the "endpoint error," i.e., the  $\infty$ -norm (see e.g., Stewart [73], pg. 164) of the difference between the true and computed solutions, evaluated at the endpoint of the interval of interest (the unit interval I). All testing was carried out on the Carnegie-Mellon University Computer Science Department's PDP-10 in ALGOL and FORTRAN, using double precision.

The first problems considered were of the form

$$(5.1) \quad \dot{x}(t) = \lambda x(t) \quad x(0) = 1$$

on the unit interval I. Although this problem is easy to handle analytically, any general problem of the form (2.1) may be locally approximated by a linear system of ordinary differential equations (see e.g., Hindmarsh [74], pp. 17-18). If the coefficient matrix of this linear system is diagonalizable, an uncoupled set of scalar equations of the form (5.1) will result.

These problems were solved via Taylor series methods; the optimal order is given in Table 5.1 for the choices of  $\lambda$  indicated. Here the optimal order was taken to be that order which minimizes the number of evaluations of the right-hand side of (5.1)

required to attain the desired error criterion. As expected, the number of evaluations required increases as the error criterion  $\epsilon$  decreases. Moreover, the optimal order also increases monotonically as  $\epsilon$  decreases, just as the theory predicts.

We next turn to the solution of the test problem

$$(5.2) \quad \dot{x}(t) = \cos^2 x(t) \quad x(0) = 0 .$$

For this problem, we searched for the optimum "unmodified" Brent-Runge-Kutta method. For this problem, the optimal order was taken to be that for which the actual CPU time (in milliseconds) required to solve the problem to within a given  $\epsilon$  was minimized. Since there is a certain amount of randomness in such a measure, the mean time for ten runs was analyzed. Not surprisingly, it turned out that the order which minimized the CPU time also minimized the number of evaluations of the right-hand side of (5.2). Since the  $(n + 2)^{\text{th}}$ -order method requires the zeros of the Jacobi polynomial  $G_n(2, 2, \cdot)$ , and the best set of values available only contained the zeros for  $1 \leq n \leq 8$  (Table 25.8 of Abramowitz and Stegun [64]), only the methods of order not exceeding ten were implemented.

The results for problem (5.2) are given in Table 5.2. Here, the optimal order  $p^*$ , the optimal number of mesh points  $n^*$ , the minimal number of evaluations  $C_e^*$ , and the minimal mean CPU time  $C_t^*$  are given. Note that these all behave as predicted. In addition, we computed the ratio of the mean CPU time for a fourth-order method  $C_t^*(4, \cdot)$  to the minimal mean runtime. As the theory predicts, this ratio appears to be increasing without bound as  $\epsilon$  tends to zero. (The same behavior was found for the ratio  $C_e(4, \cdot) / C_e^*$ , where  $C_e(4, \cdot)$  is the number of evaluations required by a fourth-order method.)

Finally, we looked at the "hard" problem

$$(5.3) \quad \dot{x}_i(t) = \sum_{j=1}^2 \alpha_{ij}(t) x_j(t) x_j(t) \quad (1 \leq i \leq 2)$$

$$\alpha_{ij}(t) = \gamma_{ij} \int_{-\infty}^{+\infty} \exp(-\nu_{ij} (t - \tau)^2) \tau^{-1} \sin \tau \, d\tau \quad (1 \leq i, j \leq 2)$$

(where "exp" denotes the exponential function), with initial conditions

$$x_1(0) = x_2(0) = 1 .$$

The  $\gamma_{ij}$  were all taken to be one, while the  $\nu_{ij}$  were taken to be

$$\nu_{11} = 1, \quad \nu_{12} = \nu_{22} = 10^{-6}, \quad \nu_{21} = 10^{-3} .$$

(This system of differential equations is similar to the system governing a two-species gas chemical reaction; see e.g., Finlayson [71].)

Since the system (5.3) is nonscalar and nonautonomous, the Brent-Runge-Kutta methods are not appropriate. Since the derivatives of  $x_i(t)$  are not readily available, the Taylor series methods are not particularly easy to apply. Thus we used linear Runge-Kutta methods for the solution of (5.3). The particular methods RKp of order p ( $1 \leq p \leq 8$ ) used were as follows.

- RK1 . . . Euler's method
- RK2 . . . Ralston [66] (5.6-40) "modified Euler"
- RK3 . . . Ralston [66] (5.6-45)
- RK4 . . . Ralston [66] (5.6-48) "classical method"
- RK5 . . . Cassity [66]
- RK6 . . . Butcher [64] (first method on page 192)
- RK7 . . . Shanks [66]
- RK8 . . . Cooper and Verner [72]

The methods of order less than eight have the optimal number of stages per step, while the method of Cooper and Verner has the minimum number of stages of all eighth-order methods known.

Most of the work involved in solving (5.3) was in evaluating  $\alpha_{ij}(t)$ . An obvious change of variable reduces this to a Gauss-Hermite quadrature; a twenty-point quadrature (Table 25.10 of Abramowitz and Stegun [64]) was used for maximal accuracy. The Chebyshev rational function approximation given on page 356 of Fröberg [69] was used to compute  $(\sin \tau) / \tau$  for  $|\tau| \leq 1$ ; the system double-precision sine routine was used for  $|\tau| > 1$ .



Since so much of the time required to solve (5.3) was spent in evaluating  $\alpha_{ij}(t)$ , the measure of cost was the number of evaluations of the set  $\{\alpha_{ij}(t) : 1 \leq i, j \leq 2\}$ ; that is, we measured the number of evaluations of the (vector) right-hand side of (5.3). (Moreover, the amount of computer time required to search for the optimum was so great as to preclude running the problem a large number of times and averaging the results, as was done in the previous example.) Results are given in Table 5.3, where  $p^*$ ,  $n^*$ , and  $C_e^*$  (defined as for (5.2)) are given as a function of the error criterion. The table stops at  $\epsilon = 10^{-5}$ , since the eighth-order method (i.e., the highest-order method implemented for testing) was reached at that level. Again, note that the theoretical results predicted are confirmed in this difficult example.

So, our three numerical examples yield data which agree with the theoretical result that the optimal order  $p^*(\alpha)$  increases with  $\alpha = \ln \epsilon^{-1}$ . Moreover, in Werschulz [76a], [76b], we show that  $p^*(\alpha) = O(\alpha)$  as  $\alpha \uparrow \infty$  for these classes of methods; i.e., the optimal order increases linearly with  $\alpha$ . The data in Tables 5.1-5.3 support this result.

TABLE 5.1

Taylor Series Methods for Test Problem

$$\dot{x}(t) = \lambda x(t) \quad x(0) = 1$$

$-\log_{10} \epsilon$	$\lambda = -e$	$\lambda = -1$	$\lambda = -1/e$	$\lambda = 1/e$	$\lambda = 1$	$\lambda = e$
1	2	3	1	1	3	8
2	9	4	2	2	4	9
3	11	6	3	3	6	11
4	12	7	4	4	7	12
5	14	8	5	5	8	14
6	15	9	6	6	9	15
7	16	10	7	7	10	16
8	17	11	8	8	11	18
9	19	12	9	9	12	19

Notes:

1. In all cases except  $\lambda = -e$ ,  $\epsilon = 10^{-1}$ , the optimal mesh-size was  $h = 1.0$ ; for this exceptional case, it was  $h = 0.5$ .
2. Entry in table is the optimal order for the given  $\lambda$  and  $\epsilon$ . This equals the minimal number of function evaluations required to solve the problem on the entire unit interval, except for the exceptional case noted above, where four was the minimal number of evaluations.

TABLE 5.2

Brent-Runge-Kutta Methods for Test Problem

$$\dot{x}(t) = \cos^2 x(t) \quad x(0) = 0$$

$-\log_{10} \epsilon$	$p^*$	$n^*$	$C_e^*$	$C_t^*$	$C_t(4, \cdot)/C_t^*$
1	1	2	2	2.789	3.93
2	2	2	4	7.024	3.28
3	4	2	6	23.144	1.00
4	5	2	8	32.401	1.38
5	6	2	10	46.837	1.87
6	7	2	12	60.970	2.15
7	8	2	14	75.613	3.18
8	9	2	16	92.052	4.58
9	10	2	18	108.632	6.85

TABLE 5.3

Linear Runge-Kutta Methods for Test Problem

$$\dot{x}(t) = \sum_{j=1}^2 \alpha_{ij}(t) x_i(t) x_j(t) \quad x_i(0) = 1 \quad (1 \leq i \leq 2)$$

$$\alpha_{ij}(t) = \gamma_{ij} \int_{-\infty}^{+\infty} \exp(-\nu_{ij} (t - \tau)^2) \tau^{-1} \sin \tau \, d\tau \quad (1 \leq i, j \leq 2)$$

$-\log_{10} \epsilon$	$p^*$	$n^*$	$C_e^*$
1	3	8	24
2	4	10	40
3	4	15	60
4	7	9	81
5	8	9	99

## Acknowledgements

I would like to thank Professors G. J. Fix, H. T. Kung, and J. F. Traub of Carnegie-Mellon University and Professor H. Woźniakowski of the University of Warsaw for their comments and suggestions on the results reported in this paper.

## References

Abramowitz and Stegun [64]:

Abramowitz, M., and I. A. Stegun, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Washington, D. C.: National Bureau of Standards, 1964.

Ahlfors [66]:

Ahlfors, L. V., Complex Analysis, Second Edition. New York: McGraw-Hill, 1966.

Brent [74]:

Brent, R. P., "Efficient Methods for Finding Zeros of Functions Whose Derivatives are Easy to Evaluate," Report, Computer Science Department, Carnegie-Mellon University, 1974.

Buck [65]:

Buck, R. C., Advanced Calculus, Second Edition. New York: McGraw-Hill, 1965.

Butcher [64]:

Butcher, J. C., "On Runge-Kutta Processes of High Order," J. Austral. Math. Soc., Vol. 4, pp. 179-194, 1964.

Cassity [66]:

Cassity, C. R., "Solutions of the Fifth-Order Runge-Kutta Equations," SIAM J. Num. Anal., Vol. 3, No. 4, pp. 598-606, December, 1966.

Cooper [69]:

Cooper, G. J., "Error Bounds for Some Single-Step Methods," Conf. on the Numerical Solution of Differential Equations, Lecture Notes in Mathematics 109, pp. 140-147. Berlin: Springer-Verlag, 1969.

Cooper and Verner [72]:

Cooper, G. J. and J. H. Verner, "Some Explicit Runge-Kutta Methods of High Order," SIAM J. Num. Anal., Vol. 9, No. 3, pp. 389-405, September, 1972.

Finlayson [71]:

Finlayson, B. E., "Covergence of the Galerkin Method for Nonlinear Problems Involving Chemical Reaction," SIAM J. Num. Anal., Vol. 8, No. 2, pp. 316-324, June, 1971.

Friedman [69]:

Friedman, A., Partial Differential Equations. New York: Holt, Rinehart, and Winston, 1969.

Fröberg [69]:

Fröberg, C. E., Introduction to Numerical Analysis, Second Edition. Reading: Addison-Wesley, 1969.

Gurtin [75]:

Gurtin, M. E., "Thermodynamics and Stability," Arch. Rat. Mech. Anal., Vol. 59, No. 1, pp. 63-96, 1975.

Henrici [62]:

Henrici, P., Discrete Variable Methods in Ordinary Differential Equations. New York: Wiley, 1962.

Hindmarsh [74]:

Hindmarsh, A. C., "Numerical Solution of Ordinary Differential Equations: Lecture Notes." Lawrence Livermore Laboratory Report No. UCID-16588, June, 1974.

Hull et al. [72]:

Hull, T. E., W. H. Enright, B. M. Fellen, and A. E. Sedgwick, "Comparing Numerical Methods for Ordinary Differential Equations," SIAM J. Num. Anal., Vol. 9, No. 4, pp. 603-637, December, 1972.

Kung and Traub [73]:

Kung, H. T. and J. F. Traub, "Computational Complexity of One-Point and Multipoint Iteration," in Complexity of Real Computation, edited by R. Karp. Providence: Amer. Math. Soc., 1973.

Kung and Traub [74]:

Kung, H. T. and J. F. Traub, "Optimal Order of One-Point and Multipoint Iteration." JACM, Vol. 21, No. 4, pp. 643-651, October, 1974.

Lindberg [74]:

Lindberg, B., "Optimal Stepsize Sequences and Requirements for the Local Error for Methods for Stiff Differential Equations." Technical Report No. 67, Department of Computer Science, University of Toronto, May, 1974.

Ralston [66]:

Ralston, A., A First Course in Numerical Analysis. New York: McGraw-Hill, 1966.

Sandberg [67]:

Sandberg, I. W., "Two Theorems on the Accuracy of Numerical Solutions of Ordinary Differential Equations," Bell Sys. Tech. J., Vol. 46, pp. 1243-1266, July-August, 1967.

Shanks [66]:

Shanks, E. B., "Solutions of Differential Equations by Evaluations of Functions," Math. Comp., Vol. 20, No. 93, pp. 21-38, January, 1966.

Stetter [73]:

Stetter, H. J., Analysis of Discretization Methods for Ordinary Differential Equations. Berlin: Springer-Verlag, 1972.

Stewart [73]:

Stewart, G. W., Introduction to Matrix Computations. New York: Academic Press, 1973.

Szegő [59]:

Szegő, G., Orthogonal Polynomials. Amer. Math. Soc. Colloquium Publications, Vol. XXIII. New York: Amer. Math. Soc., 1959.

Traub [64]:

Traub, J. F., Iterative Methods for the Solution of Equations. Englewood Cliffs: Prentice-Hall, 1964.

Traub [72]:

Traub, J. F., "Computational Complexity of Iterative Processes," SIAM J. Comput., Vol. 1, No. 2, pp. 167-179, June, 1972.

Traub and Woźniakowski [76]:

Traub, J. F. and H. Woźniakowski, "Strict Lower and Upper Bounds on Iterative Complexity," in Analytic Computational Complexity, edited by J. F. Traub. New York: Academic Press, 1976.

Werschulz [76a]:

Werschulz, A. G., "Computational Complexity of One-Step Methods for Systems of Differential Equations," Report, Computer Science Department, Carnegie-Mellon University, 1976.

Werschulz [76b]:

Werschulz, A. G., "Computational Complexity of One-Step Methods for a Scalar Autonomous Differential Equation," Report, Computer Science Department, Carnegie-Mellon University, 1976.

Woźniakowski [75]:

Woźniakowski, H., "Generalized Information and Maximal Order of Iteration for Operator Equations," SIAM J. Num. Anal., Vol. 12, No. 1, pp. 121-135, March, 1975.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS  
BEFORE COMPLETING FORM

1. REPORT NUMBER  
2. GOVT ACCESSION NO. 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle)  
OPTIMAL ORDER AND MINIMAL COMPLEXITY OF ONE-STEP METHODS FOR INITIAL VALUE PROBLEMS
5. TYPE OF REPORT & PERIOD COVERED  
Interim
6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s)  
Arthur G. Werschulz
8. CONTRACT OR GRANT NUMBER(s)  
N00014-76-C-0370,  
NR 044-422
9. PERFORMING ORGANIZATION NAME AND ADDRESS  
Carnegie-Mellon University  
Computer Science Dept.  
Pittsburgh, PA 15213
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS  
Office of Naval Research  
Arlington, VA 22217
12. REPORT DATE  
September 1976
13. NUMBER OF PAGES  
39
14. MONITORING AGENCY NAME & ADDRESS (// different from Controlling Office)
15. SECURITY CLASS, (of this report)

UNCLASSIFIED  
15a. DECLASSIFICATION/ DOWNGRADING  
SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

*Abstract:* We consider the task of numerically approximating the solution of an ordinary differential equation initial value problem. A methodology is given for determining the computational complexity of finding an approximate solution with error not exceeding  $c$ . In addition, we determine the method of optimal order within a given class of methods, and show that under reasonable hypotheses, the optimal order increases as  $c$  decreases, tending to infinity as  $s$  tends to zero.