MEMORANDUM
RM-2924-PR
DECEMBER 1961

SOME ASPECTS OF QUASILINEARIZATION

Robert Kalaba

PREPARED FOR:
UNITED STATES AIR FORCE PROJECT RAND

The RAND Corporation
SANTA MONICA • CALIFORNIA

MEMORANDUM

RM-2924-PR

December 1961

SOME ASPECTS OF QUASILINEARIZATION

Robert Kalaba

This research is sponsored by the United States Air Force under Project RAND — Contract No. AF 49(638)-700 — monitored by the Directorate of Development Planning, Deputy Chief of Staff, Research and Technology, Hq USAF. Views or conclusions contained in this Memorandum should not be interpreted as representing the official opinion or policy of the United States Air Force.

The RAND Corporation

1700 MAIN ST. • SANTA MONICA • CALIFORNIA

PREFACE

This paper was prepared by invitation for publication in the Proceedings of the International Symposium on Nonlinear Differential Equations and Nonlinear Mechanics, held July 31 to August 4, 1961 at the Air Force Academy, Colorado Springs, Colorado, and sponsored in part by the United States Air Force. The material presented here represents part of RAND's continuing study of means for exploiting more fully the capabilities of modern computing machines in solving problems of technology and physics.

SUMMARY

Many of the fundamental nonlinear differential equations of mathematical physics and engineering can be written in the form

$$L_1(u) = \max_q \left\{ L_2(u,x,q) + a(x,q) \right\}$$

where L_1 and L_2 are linear differential operators on u, a scalar function of the vector x, and q is a decision variable. Among these equations are the Riccati equation, which plays a role in studies of wave propagation, neutron transport, and filter theory; the Emden-Fowler equation, of importance in astrophysical and nuclear studies; the Hamilton-Jacobi equation of mechanics; the eikonal equation of optics; and others. In addition, it is a basic equation of dynamic programming.

In this paper a formula giving a representation for the solution of the above type of equation is presented. It involves use of "max" operators applied to solutions of associated linear equations. In turn, this representation formula leads to the construction of quadratically convergent and monotonic sequences of functions which are of utility in the computational solution of nonlinear boundary value problems. Results of some numerical experiments involving both ordinary and partial differential equations are presented.

I. INTRODUCTION TO QUASILINEARIZATION

Early in the development of the theory of dynamic programming it was recognized that functional equations of the form

$$L_1(u) = \max_q \left\{ L_2(u,x,q) + a(x,q) \right\} \tag{1}$$

where L_1 and L_2 are linear operators on u, a scalar function of the vector x, and q is a decision vector, an element of a space Q, would play a significant role [1]. In 1955 R. Bellman pointed out that many of the classical nonlinear equations of mathematical physics and engineering could be reduced to the form of Eq. (1). The list of such equations is now known to be very large and includes the Riccati equation [2,3], the eikonal equation of optics [4], various forms of the Hamilton-Jacobi equation of mechanics [5,6], and the Emden-Fowler equation of astrophysics and nuclear physics [7], as well as various nonlinear diffusion and potential equations [4].

A detailed discussion of a variety of special cases of Eq. (1), and of the general equation itself is given in [3]. Among the equations discussed are the initial value problem

$$u' = f(u,x), \quad u(0) = c \tag{2}$$

the boundary value problem

$$u'' = f(u,x), \quad u(0) = u(b) = 0 \tag{3}$$

and the nonlinear potential and diffusion equations

$$(4) \qquad u_{xx} + u_{yy} = f(u,x,y)$$

$$(5) \qquad u_t - u_{xx} = f(u,x,y)$$

Let us drop the maximum operator in Eq. (1) and consider the associated linear equation for $w(x,q)$

$$(6) \qquad L_1(w) = L_2(w,x,q) + a(x,q)$$

Then, under certain circumstances, the solution of the original Eq. (1) may be represented in the form

$$(7) \qquad u(x) = \max_q w(x,q)$$

This is the essence of quasilinearization. The representation in Eq. (7) leads to the production of lower bounds for the solution of Eq. (1). In addition, it leads to a method of successive approximations for the solution of nonlinear boundary value problems, the convergence being both monotone and quadratic. Consequently the technique, abstractly equivalent to the Newton-Raphson method, is of interest in the computational solution of these equations.

The aim of this paper is to provide a sketch of some of the principal ideas and applications associated with the notion of quasilinearization, and is intended to be quite self-contained. We shall first show how equations of the form of Eq. (1) arise from application of the principle of optimality to a simple problem in the calculus of variations. Next we examine the Riccati equation. Section IV is devoted to a discussion of the nonlinear boundary value problem arising from minimization of the integral

$$\int_0^1 \left\{ke^y + \frac{1}{2}(y'')^2\right\} dx$$

subject to the conditions

$$y(0) = y'(0) = 0$$

$$y \text{ at } x = 1 \text{ is free}$$

In particular, some interesting numerical results are presented since one of the major applications is to the solution of nonlinear boundary value problems, many of which arise from variational contexts.

Section V contains some remarks on the computational solution of boundary value problems for nonlinear partial differential equations. Finally, a discussion of some additional matters is given.

II. A VARIATIONAL PROBLEM

Equations of the form of Eq. (1) occur naturally in analysis. To show this we consider a classical variational problem from the functional equation viewpoint of dynamic programming.

We consider the problem of maximizing the integral

$$(8) \qquad \int_0^T F(x,y)dt = J[y]$$

where

$$(9) \qquad \frac{dx}{dt} = G(x,y), \quad x(0) = c$$

Many problems in the theory of automatic control are of this type. First we observe that the maximal value is solely a function of c, the initial state, and T, the duration of the process,

$$(10) \qquad \max_{y(t)} J[y] = f(c,T)$$

Next we use Bellman's principle of optimality [1] to see that

$$(11) \qquad f(c,T) = \max_{Y} \left\{ F(c,Y)\Delta + f(c + G(c,Y)\Delta, T-\Delta) + o(\Delta) \right\}$$

where $Y = y(0)$. Of course, the maximizing value of Y depends on c and T. Then, under suitable regularity assumptions, a passage to the limit yields the nonlinear differential equation

(12) $$f_T = \max_Y \left\{ F(c,Y) + G(c,Y)f_c \right\}$$

Furthermore, as an initial condition, we have

(13) $$f(c,0) = 0$$

We assume that these equations have a unique solution in some region for which $T \geq 0$. Equation (12) is of the form which we wish to study.

Next let us consider the linear differential equation which we obtain by dropping the maximum operator in Eq. (12):

(14) $$w_T = F(c,Y) + G(c,Y)w_c$$

(15) $$w(c,0) = 0$$

We regard Y as a given function of c and T. In view of Eqs. (12) and (14) we see that

(16) $$(f - w)_T = G(c,Y)\ (f - w)_c + p$$

where

(17) $$p \geq 0$$

In addition for $T = 0$ we find

(18) $$f(c,0) - w(c,0) = 0$$

But Eqs. (16) and (17) imply that for some region for which $T \geq 0$

$$f(c,T) - w(c,T) \geq 0 \tag{19}$$

$$f(c,T) \geq w(c,T) \tag{20}$$

On the other hand, equality will hold for an appropriate choice of the function Y in Eq. (14) which determines the function w. This leads to the representation

$$f(c,T) = \max_{Y} w \tag{21}$$

We have now represented the solution of the nonlinear Eq. (12) in terms of a maximization over Y of solutions of the associated linear Eq. (14).

The purpose of this section has been to show that equations of the form of Eq. (1) occur naturally in analysis. Rather than continue along these lines, let us now turn to a study of one of the classical equations, that of Riccati, to illustrate our methodology.

III. THE RICCATI EQUATION AND A GENERALIZATION

Let us consider the Riccati equation

$$\frac{du}{dx} = u^2 + g(x), \quad 0 \leq x \leq b \tag{22}$$

$$u(0) = c \tag{23}$$

where $g(x)$ is continuous, and b is sufficiently small to guarantee the existence of a solution. Since

$$u^2 \geq 2vu - v^2 \tag{24}$$

with equality holding for

$$v \equiv u \tag{25}$$

we may rewrite Eq. (22) in the form

$$\frac{du}{dx} = \max_{v} \left\{ 2vu - v^2 + g \right\} \tag{26}$$

This shows that the Riccati equation can be transformed into an equation having the form of Eq. (1).

To proceed with the analysis we introduce the associated linear equation

$$\frac{dw}{dx} = 2vw - v^2 + g, \quad w(0) = c \tag{27}$$

which we obtain by dropping the maximum operator in Eq. (26). To indicate that w depends on x and the choice of the function $v = v(x)$ we write

(28) $$w = w[x;v]$$

Our first objective is to show that u may be represented in the form

(29) $$u = \max_{v} w[x;v]$$

where the maximization is over all functions $v = v(x)$, $0 \leq x \leq b$ for which the integrals in Eq. (36) below exist.

Equation (26) may be written

(30) $$\frac{du}{dx} = 2vu - v^2 + p + g$$

where p is a non-negative function on the interval [0,b]. It follows that the difference function $z = u - w$ satisfies the equation

(31) $$\frac{dz}{dx} = 2vz + p, \quad z(0) = 0$$

for which the solution is

(32) $$z(x) = \int_0^x p(s)\exp\left\{2\int_s^x v(t)dt\right\} ds$$

The non-negativity of this integral shows that

(33) $$z(x) \geq 0, \quad 0 \leq x \leq b$$

In addition we know that if we select

(34) $$v(x) = u(x)$$

then the solution of Eq. (27) is

$$(35) \qquad w[x;u(x)] = u(x)$$

This completes the proof of the representation of the solution of the Riccati Eq. (22) in the form of Eq. (29). By solving Eq. (27) for w we find the interesting representation for the solution of the Riccati equation

$$(36) \qquad u(x) = \max_{v} \left\{ c \exp\left[2 \int_0^x v(s)ds \right] + \int_0^x [g(s) - v^2(s)] \cdot \exp\left\{ 2 \int_s^x v(t)dt \right\} ds \right\}, \quad 0 \leq x \leq b$$

It follows, of course, that uniform lower bounds for u can be obtained by evaluating the right-hand side of Eq. (36) for simple admissible functions $v = v(x)$.

As a slight digression, let us note that we can use the representation formula just derived to obtain a representation for the solution of the linear homogeneous equation of second order

$$(37) \qquad \frac{d^2y}{dx^2} + g(x)y = 0, \quad 0 \leq x \leq b$$

$$(38) \qquad y(0) = y_0 \neq 0, \quad y'(0) = y_1$$

We use the well-known transformation

$$(39) \qquad -\frac{y'}{y} = u$$

to rewrite Eq. (37) as

$$(40) \qquad \frac{du}{dx} = u^2 + g(x)$$

$$(41) \qquad u(0) = -y_1/y_0 = c$$

Then application of Eq. (36) yields the result

$$(42) \qquad \frac{y(x)}{y_0} = \min_v \exp\left\{ - \int_0^x \left[c \exp\left\{ 2\int_0^s v(t)dt \right\} + \int_0^s (g(t) - v^2(t))\exp\left\{ 2\int_t^s v(t_1)dt_1 \right\} dt \right] ds \right\}$$

$$0 \le x \le b$$

for b sufficiently small.

We now turn to the problem of constructing monotone sequences of functions which converge to u, the solution of Eq. (22). Choose an arbitrary initial function $v = v_0(x)$ and determine the solution of the associated linear Eq. (27),

$$(43) \qquad \frac{du_0}{dx} = 2v_0u_0 - v_0^2 + g, \quad u_0(0) = c$$

Then determine an improved choice of the function v, call it v_1, as the function which maximizes the expression

$$2u_0v - v^2 + g$$

This maximizing function is

(44) $$v = v_1 = u_0$$

Next, compute the function $u_1(x)$ as the solution of the equation

(45) $$\frac{du_1}{dx} = 2u_0u_1 - u_0^2 + g, \quad u_1(0) = 0$$

Then determine an improved choice of the function v, $v = v_2(x)$, as the function which maximizes

$$2u_1v - v^2 + g$$

This maximizing function is

(46) $$v = v_2 = u_1$$

Continuing in this way we are led to consider the sequence of functions $\{u_n\}$ defined by the linear equations

(47) $$\begin{cases} \dfrac{du_0}{dx} = 2v_0u_0 - v_0^2 + g, & u_0(0) = c \\[2ex] \dfrac{du_{n+1}}{dx} = 2u_nu_{n+1} - u_n^2 + g, & u_{n+1}(0) = c, \quad n \geq 0 \end{cases}$$

This type of reasoning is familiar in dynamic programming and is known as approximation in policy space [1].

The recurrence relations of Eq. (47) may be viewed as an extension of Newton's method, for the right-hand side of the latter equations may be considered to arise from expanding $f(u) = u^2 + g$ around $u = u_n$ and keeping only the linear terms, $(u_n^2 + g) + (u_{n+1} - u_n)2u_n$. In the more familiar Picard scheme of successive approximations only the first term, $u_n^2 + g$, is retained. As we shall see, in some cases the Newtonian scheme has marked computational advantages.

The sequence of functions $u_n(x)$ tends monotonically to $u(x)$, and, in addition, the convergence is quadratic.

Consider the equations

$$(48) \qquad \frac{du_{k+1}}{dx} = 2u_k u_{k+1} - u_k^2 + g, \quad u_{k+1}(0) = c$$

$$(49) \qquad \frac{du_k}{dx} = 2u_{k-1} u_k - u_{k-1}^2 + g, \quad u_k(0) = c, \ k = 0,1,\ldots$$

The maximum of the right-hand side of Eq. (49) is assumed when $u_{k-1} = u_k$, which implies that

$$(50) \qquad \frac{du_k}{dx} \leq 2u_k^2 - u_k^2 + g$$

Consequently, we have

$$(51) \qquad d(u_{k+1} - u_k)/dx \geq 2u_k(u_{k+1} - u_k), \quad u_{k+1}(0) - u_k(0) = 0$$

which, in view of Eqs. (31) and (33), implies that

$$(52) \qquad u_{k+1}(x) \geq u_k(x)$$

where

$$0 \leq x \leq b$$

This is the desired monotonicity result.

To demonstrate the quadratic nature of the convergence, which is of great importance from the computational viewpoint, we note that

$$(53) \qquad \frac{d(u-u_n)}{dx} = u^2 - \left[u_{n-1}^{\;2} + 2u_{n-1}(u_n - u_{n-1})\right]$$

This can be rewritten to yield

$$(54) \qquad \frac{d(u-u_n)}{dx} = u^2 - u_{n-1}^{\;2} - (u-u_{n-1})2u_{n-1} - (u_n-u)2u_{n-1}$$

Since

$$(55) \qquad u(0) - u_n(0) = 0$$

upon integrating from 0 to x, we find

$$(56) \qquad u(x) - u_n(x) = \int_0^x \left[(u - u_{n-1})^2 + (u - u_n)2u_{n-1}\right]ds$$

$$0 \le x \le b$$

Upon introducing the norm

$$(57) \qquad \|u - u_n\| = \max_{0\le x\le b} |u(x) - u_n(x)|$$

we find

$$(58) \qquad \|u - u_n\| \le b\left\{ \|u - u_{n-1}\|^2 + \|u - u_n\| \, m\right\}$$

where

$$(59) \qquad \left|2u_{n-1}\right| \leq m, \quad 0 \leq x \leq b$$

(A region of uniform boundedness for the sequence $\{u_n\}$ is readily established.) Thus

$$(60) \qquad \left\|u - u_n\right\| \leq \frac{b}{1 - bm} \left\|u - u_{n-1}\right\|^2$$

which demonstrates the quadratic convergence on a sufficiently small interval [0,b].

Similar considerations can be carried out for a generalization of the Riccati equation

$$(61) \qquad u' = f(u,x), \quad u(0) = c$$

under the assumption that f is a convex function of u for all x [3]. For in this case we can use the well-known property of convex functions that

$$(62) \qquad f(u,x) = \max_v \left\{ f(v,x) + (u - v)\frac{\partial f}{\partial v} \right\}$$

Thus Eq. (61) can be written

$$(63) \qquad u' = \max_v \left\{ f(v,x) + (u - v)\frac{\partial f}{\partial f} \right\}$$

a form for which our formalism is applicable. See [3] for details.

IV. A NUMERICAL EXPERIMENT WITH A NONLINEAR BOUNDARY VALUE PROBLEM

Consider the following variational problem. Determine the function $y = y(x)$ on the interval $[0,1]$ so as to minimize the integral

$$(64) \qquad J[y] = \int_0^1 \left\{ ke^y + \frac{1}{2}(y'')^2 \right\} dx$$

We assume that at $x = 0$ the function y satisfies the conditions

$$(65) \qquad y(0) = y'(0) = 0$$

and at $x = 1$ it satisfies the natural boundary conditions

$$(66) \qquad y''(1) = y'''(1) = 0$$

The Euler equation for this problem is

$$(67) \qquad y^{IV} + ke^y = 0$$

Consequently we wish to find an efficient computational procedure for solving the nonlinear Eq. (67) subject to the boundary conditions Eqs. (65) and (66). Our aim is to convert this boundary value problem into a rapidly converging sequence of initial value problems.

The procedure which we shall use is the following. We first select an initial approximation $v_0(x) \equiv 0$. Then the higher approximations are determined by solving the linear recurrence relations

$$(68) \qquad u_0^{IV} = -ke^{v_0} - k(u_0 - v_0)e^{v_0}$$

(69) $$u_{n+1}^{IV} = -ke^{u_n} - k(u_{n+1} - u_n)e^{u_n}$$

the boundary conditions being

(70) $$u_k(0) = u_k'(0) = 0$$

$$u_k''(1) = u_k'''(1) = 0, \quad k = 0,1,2,\ldots$$

The computational determination of the function $u_{n+1}(x)$, given $u_n(x)$, is quite simple. First the particular solution of Eq. (69), $p(x)$, subject to the initial conditions $p(0) = p'(0) = p''(0) = p'''(0) = 0$ is computed and stored. Then the homogeneous solution, $w(x)$, satisfying the initial conditions $w(0) = w'(0) = w''(0) = 0$, $w'''(0) = 1$, is computed and stored. Next the homogeneous solution, $z(x)$, satisfying the initial conditions $z(0) = z'(0) = z'''(0) = 0$, $z''(0) = 1$ is computed and stored. Consequently the desired function $u_{n+1}(x)$ can be expressed in the form

(71) $$u_{n+1}(x) = p(x) + Aw(x) + Bz(x)$$

The initial conditions $u_{n+1}(0) = u_{n+1}'(0) = 0$ are automatically satisfied. The constants A and B are determined by using the computed values of $p(1)$, $w(1)$ and $z(1)$ in conjunction with the remaining conditions

(72) $$u_{n+1}''(1) = u_{n+1}'''(1) = 0$$

This required merely the solution of two linear equations in two unknowns. This completes the computational determination of $u_{n+1}(x)$. Notice that only the solution of initial value problems is required, along with the solution of some linear algebraic equations.

Now let us look at some computational results for the case $k = 6$, showing the rapidity of convergence. The entire calculation consumed only a few seconds on the IBM 7090. An Adams-Moulton integration procedure yielding about six significant figures was employed.*

Table 1

SOME COMPUTATIONAL RESULTS $(v_0(x) = 0)$

x	$u_0(x)$	$u_1(x)$	$u_2(x)$	$u_3(x)$
0.0	0.0	0.0	0.0	0.0
0.20312	-0.375572×10^{-1}	-0.394922×10^{-1}	-0.394976×10^{-1}	-0.394976×10^{-1}
0.40625	-0.128544	-0.135598	-0.135617	-0.135617
0.60156	-0.242775	-0.256792	-0.256831	-0.256831
0.8469	-0.373308	-0.395724	-0.395786	-0.395786
1.0	-0.501782	-0.532645	-0.532730	-0.532730

When k becomes sufficiently negative the minimum problem can fail to have a solution, and, of course, the scheme breaks down. Apparently this happens for k lying between - 4 and - 5.

Similar techniques have been used successfully by Collatz [8], Hestenes [9], and others.

---

*The programming was very capably handled by Dr. Bella Kotkin, of RAND.

In applying the quasilinearization technique to higher order equations or to systems of equations, especially where high accuracy is required, problems due to the limited high-speed storage of modern computers (about 32,000 words, at best) arise. A powerful technique for avoiding them is discussed in the paper by Bellman [10].

V. COMPUTATIONAL SOLUTION OF SOME NONLINEAR PARTIAL DIFFERENTIAL EQUATIONS

In Part III of [3] treatments of both the nonlinear potential equation

$$\Delta u = f(u) \tag{73}$$

and the nonlinear diffusion equation

$$u_t - u_{xx} = f(u) \tag{74}$$

under various boundary conditions, are given using the quasilinearization technique. In [9] a report is given concerning the results of numerical experiments which were conducted to compare the relative merits of the Picard approximation scheme versus the Newtonian scheme. The advantages of a quadratically convergent scheme over a linearly convergent scheme are not so clear in the case of partial differential equations. There are many reasons for this, but among the most prominent is the difficulty in solving the associated linear equations to a sufficiently high degree of accuracy. Nevertheless, we did find that in those cases for which the solution has large gradients, the higher order method was superior.

VI. DISCUSSION

Though the wide applicability of the notion of quasilinearization is by now apparent, let us mention several additional ones which are discussed in [11,3]. In the first place, systems of equations can be treated, provided that an inequality between vectors implies that relation between components. In addition some of P. Lax's theory of weak solutions of nonlinear first-order partial differential equations can be derived [12]. Also various nonlinear integral and differential-integral equations are amenable to treatment. Much remains to be done.

REFERENCES

1. Bellman, R., Dynamic Programming, Princeton University Press, Princeton, N. J., 1957.

2. Bellman, R., "Functional Equations in the Theory of Dynamic Programming. II. Nonlinear Differential Equations," Proc. Nat. Acad. Sci. USA, Vol. 41, 1955, pp. 482-485.

3. Kalaba, R., "On Nonlinear Differential Equations, the Maximum Operation, and Monotone Convergence," Jour. Math. and Mech., Vol. 8, 1959, pp. 519-574.

4. Kalaba, R., "Dynamic Programming, Fermat's Principle, and the Eikonal Equation," J. Optical Soc. Amer., Vol. 51, 1961, pp. 1150-1151.

5. Kalaba, R., "Dynamic Programming and the Variational Principles of Classical and Statistical Mechanics," Proc. Fifth Midwestern Conference on Solid Mechanics, Plenum Press, New York, 1961.

6. Dreyfus, S., "Dynamic Programming and the Calculus of Variations," J. Math. Anal. and Appl., Vol. 1, 1960, pp. 228-239.

7. Bellman, R., Stability Theory of Differential Equations, McGraw-Hill Book Co., New York, 1953.

8. Collatz, L., "Monotonie und Extremalprinzipien beim Newtonschen Verfahren," Numerische Mathematik, Vol. 3, 1961, pp. 99-106.

9. Bellman, R., M. Juncosa, and R. Kalaba, "Some Numerical Experiments Using Newton's Method for Nonlinear Parabolic and Elliptic Boundary-Value Problems," Comm. Assoc. Comput. Machinery, Vol. 4, 1961, pp. 187-191.

10. Bellman, R., "Successive Approximations and Computer Storage Problems in Ordinary Differential Equations," Comm. Assoc. Comput. Machinery, Vol. 4, 1961, pp. 222-223.

11. Bellman, R., "Functional Equations in the Theory of Dynamic Programming. V. Positivity and Quasi-Linearity," Proc. Nat. Acad. Sci. USA, Vol. 41, 1955, pp. 743-748.

12. Lax, P., Initial Value Problems for Nonlinear Hyperbolic Equations, University of Kansas, Technical Report 14, 1955.

ADDITIONAL REFERENCE

Kalaba, R., "Computational Considerations for Some Deterministic and Adaptive Control Processes," Optimization Techniques, Academic Press, New York, 1962.