COMPUTATIONAL EXPERIENCE WITH THE CHOW-YORKE ALGORITHM

by

Layne T. Watson

CS78010-R

November 1978

Department of Computer Science
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061

This work was partially supported by NSF Grant MCS 7821337.

Abstract. The Chow-Yorke algorithm is a nonsimplicial homotopy-type method for computing Brouwer fixed points that is globally convergent. It is efficient and accurate for fixed point problems. L. T. Watson, T. Y. Li, and C. Y. Wang have adapted the method for zero finding problems, the nonlinear complementarity problem, and nonlinear two-point boundary value problems. Here theoretical justification is given for applying the method to some mathematical programming problems, and computational results are presented.

<u>Introduction</u>. The Chow-Yorke algorithm was proposed in abstract terms in [3], and developed into a numerical algorithm by Watson [20]. The algorithm was originally conceived for computing Brouwer fixed points, but has been adapted for zero finding problems [3,20], the nonlinear complementarity problem [21], nonlinear two-point boundary value problems [17, 18, 23], and some continuum mechanics problems [19]. The algorithm requires C^2 differentiability, but that has not been a serious restriction so far. Considerable effort has gone into developing good computer code for the algorithm, and a code using very sophisticated mathematical software is in [22].

The numerical algorithm is described in detail in [20], so only a brief sketch will be given here. Let $f:B\to B$ be a C^2 map, where B is the closed unit ball (or any compact, convex subset) in n-dimensional Euclidean space E^n. Define the homotopy map $\rho_a:[0,1) \times B\to E^n$ by

$$\rho_a(\lambda,x) = \lambda(x-f(x))+(1-\lambda)(x-a).$$

There is a theorem from differential geometry (known as a parameterized Sard's theorem) which states that for almost all a in the interior of B, ρ_a is transversal to zero (i.e., the Jacobian matrix $D\rho_a(\lambda,x)$ has full rank on $\rho_a^{-1}(0)$). This result is discussed further in [3], [20], [21], or [23]. The result can be paraphrased as: for almost all a in the interior of B there is a zero curve γ of ρ_a , along which the Jacobian matrix $D\rho_a$ has full rank, emanating from (0,a) and reaching a fixed point $\bar{x}$ of f (at $\lambda=1$). The Chow-Yorke algorithm is beautifully simple: follow the zero curve γ of ρ_a from $(0,a)$ to a fixed point $\bar{x}$ of f. The algorithm is globally convergent with probability one, in the sense that it works for all starting points a, except possibly those in a set of Lebesgue measure zero. The homotopy map $\rho_a(\lambda,x)$ resembles well-known continuation methods [2, 4, 14], but is fundamentally different. In standard continuation methods the imbedding parameter λ is an independent variable which increases monotonically from zero to one. Such methods fail if the zero curve γ turns back, because the Jacobian matrix $D_x\rho_a(\lambda,x)$ is singular at a turning point. In the Chow-Yorke algorithm, λ is a dependent variable and arc length is used as the independent variable. Since the Jacobian matrix $D\rho_a(\lambda,x)$ has full rank along γ, turning points on γ cause no difficulty whatsoever [20]. The power of the Chow-Yorke algorithm derives from this ability of λ to both increase and decrease along γ. γ is followed by solving an ordinary differential equation; see [20] for the details.

To find a zero of the C^2 map $F:E^n \to E^n$, the homotopy map would be

$$\psi_a(\lambda,x)=\lambda F(x)+(1-\lambda)(x-a).$$

Now, though, the zero curve γ may wander off to infinity, and assumptions on F are needed to guarantee that γ reaches a zero of F. Roughly speaking, these conditions of F are the same as those needed to guarantee the existence of a zero of F. For other types of problems (continuum mechanics [19], fluid dynamics [17, 18], two-point boundary value [23]) the form of the homotopy map may be different, but the essential features are getting the homotopy map transversal to zero and proving that the zero curve γ reaches a solution.

Another approach to computing fixed points is the simplicial approximation technique pioneered by Scarf [16], and refined by Eaves [5, 6] and Merrill [12]. Simplicial approximation algorithms have been extensively developed by Saigal [15] and Gould [8]. The theory and implementation of simplicial approximation and n-dimensional triangulation is relatively complex, but Saigal [15] has developed a simplicial homotopy algorithm that is competitive with the Chow-Yorke algorithm. Another homotopy algorithm similar in spirit to the Chow-Yorke algorithm is that of Kellogg, Li, and Yorke [9]. This latter method has its basis in differential geometry rather than topology, and also tracks a zero curve of some homotopy map. However, the numerical properties of the Chow-Yorke algorithm are much better than the Kellogg-Li-Yorke algorithm [20].

The intent of this paper is merely to justify the application of the Chow-Yorke algorithm to mathematical programming problems, and give some computational results. No serious attempt is made to compare different fixed point algorithms. The author agrees completely with a remark by L. F. Shampine that "...how a method is implemented may be more important than the method itself." The numerical results given here were obtained using the computer code in [22], and the superior performance of the Chow-Yorke algorithm (compared to [15], for example) may be due to the sophisticated implementation rather than the algorithm itself.

Theoretical justification. The theoretical foundation of the Chow-Yorke algorithm is developed in [3] and elaborated on in [20]. For completeness here, the necessary background will be summarized in two lemmas. Let E^n denote n-dimensional Euclidean space.

Lemma 1. Let $F:E^n \to E^n$ be a C^2 map. For almost all $a \in E^n$, the map $\rho_a:[0,1) \times E^n \to E^n$ defined by

$$\rho_a(\lambda,x)=\lambda F(x)+(1-\lambda)(x-a)$$

is transversal to zero (i.e., for almost all a the Jacobian matrix $D\rho_a$ has full rank on $\rho_a^{-1}(0)$).

Lemma 2. Let $F:E^n \to E^n$ be a C^2 map. Then for almost all $a \in E^n$, there exists a zero curve γ of $\rho_a(\lambda,x)$ emanating from $(0,a)$ along which the Jacobian matrix $D\rho_a(\lambda,x)$ has full rank. γ either reaches a zero $\bar{x}$ of $F(x)$ (at $\lambda=1$) or wanders off to infinity. Furthermore, if γ reaches $\bar{x}$ and the Jacobian matrix $DF(\bar{x})$ is nonsingular, then γ has finite arc length.

Versions of Lemmas 1 and 2 are proved and discussed at length in [3], [20], [21], and [23]. Lemma 1 is known as a "parameterized Sard's theorem". Lemma 2 merely restates Lemma 1 in a form more useful in the present context. Also note that γ is a C^1 curve which does not intersect itself.

Theorem 1. Let $f:E^n \to E$ be a C^3 convex map with a minimum at $\tilde{x}$, $\|\tilde{x}\| \leq M$. Then for almost all a, $\|a\| < M$, there is a zero curve γ of the homotopy map

$$\rho_a(\lambda,x)=\lambda\nabla f(x)+(1-\lambda)(x-a),$$

along which the Jacobian matrix $D\rho_a(\lambda,x)$ has full rank, connecting $(0,a)$ to $(1,\bar{x})$, where $f(\bar{x})=\min_x f(x)$. Furthermore, if the Hessian matrix $H(\bar{x})=D(\nabla f)(\bar{x})$ is nonsingular, γ has finite arc length.

Proof. The existence of γ and full rank of $D\rho_a(\lambda,x)$ along γ for almost all a follow from Lemma 2. Again by Lemma 2, to prove that γ reaches a zero of ∇f (i.e., a minimum of f), it is sufficient to prove that γ is bounded. Let (λ,x) be any point, $0 \leq \lambda < 1$, $\|x\| = 3M$. Since $\|a\| < M$ and $\|\tilde{x}\| \leq M$,

$$(x-\tilde{x})(x-a)>0 \quad .$$

Since f is convex and $\tilde{x}$ is a minimum [11],

$$(x-\tilde{x})\nabla f(x)=(x-\tilde{x})(\nabla f(x)-\nabla f(\tilde{x}))\geq 0 \; .$$

Therefore

$$(x-\tilde{x})[\lambda\nabla f(x)+(1-\lambda)(x-a)]>0 \; ,$$

which implies that $\rho_a(\lambda,x)\neq 0$ for $0\leq\lambda<1$ and $\|x\| = 3M$. Thus γ is contained in $[0,1] \times \{x \mid \|x\| \leq 3M\}$, hence bounded. The finite arc length statement also follows from Lemma 2. Q.E.D.

Let $f:E^n \to E$ be a C^3 convex function. This degree of differentiability is necessary because the homotopy map ρ_a in Lemmas 1 and 2 must be C^2. By Theorems 7.2.1 and 7.3.7 in Mangasarian [11], the constrained optimization problem

(1) $$\min f(x) \quad \text{such that} \quad x \geqq 0$$

is equivalent to the nonlinear complementarity problem

(2) $$x \geqq 0, \quad F(x) \geqq 0, \quad xF(x)=0,$$

where $F(x)=\nabla f(x)$. It was shown in [10] that the nonlinear complementarity problem (2) is equivalent to

(3) $$K(x)=0,$$

where

$$K_i(x)=-|F_i(x)-x_i|^3 + (F_i(x))^3 + x_i^3 \quad .$$

Now take

$$\rho_a(\lambda,x)=\lambda K(x) + (1-\lambda)(x-a) \quad .$$

Remark 3 of [21] gives

<u>Theorem 2</u>. Suppose every zero of $K(x)$ lies in the ball $\|x\|<r$, where r is such that $x \geqq 0$ and $\|x\| \geqq r$ imply $x_k>0$ and $F_k(x) \geqq 0$ for some index k. Then there exists $\delta>0$ such that for almost all $a \geqq 0$ with $\|a\|<\delta$ there is a zero curve γ of $\rho_a(\lambda,x)$, along which $D\rho_a(\lambda,x)$ has full rank, connecting $(0,a)$ to $(1,\bar{x})$, where $\bar{x}$ is a zero of $K(x)$.

<u>Def</u>. $f:E^n \to E$ is uniformly convex if f is C^2 and there exists $\nu>0$ such that $x[H(z)]x \geqq \nu\|x\|^2$ for all z and x, where $H(z)$ is the Hessian matrix of f at z.

Theorem 3. Let $f:E^n \to E$ be a C^3 uniformly convex map. Then there exists $\delta > 0$ such that for almost all $a \geq 0$ with $\|a\| < \delta$ there is a zero curve γ of

$$\rho_a(\lambda,x) = \lambda K(x) + (1-\lambda)(x-a),$$

where

$$K_i(x) = -\left|\frac{\partial f}{\partial x_i}(x) - x_i\right|^3 + \left(\frac{\partial f(x)}{\partial x_i}\right)^3 + x_i^{\,3} ,$$

along which $D\rho_a(\lambda,x)$ has full rank, connecting $(0,a)$ to $(1,\bar{x})$, where $\bar{x}$ solves (1).

Proof. Since (1), (2), and (3) are equivalent, it suffices to prove that $F(x) = \nabla f(x)$ satisfies the hypotheses of Theorem 2. A uniformly convex function is also strictly convex, hence (1) has at most one solution, and all zeros (if there are any) of K lie in some ball $\|x\| < r$. Now $x\nabla f(x) = x[\nabla f(0) + H(z)x] = x\nabla f(0) + xH(z)x \geq \nu\|x\|^2 - \|x\| \; \|\nabla f(0)\| > 0$ for $r \leq \|x\|$ sufficiently large. Therefore $x \geq 0$, $\|x\| \geq r$ imply $x_k > 0$ and $(\nabla f(x))_k \geq 0$ for some index k. The conclusion now follows from Theorem 2. Q.E.D.

Theorems 1 and 3 guarantee that the Chow-Yorke algorithm is globally convergent with probability one for convex unconstrained minimization problems and convex problems with nonnegativity constraints. This is not a wide class of problems, but it is a good starting point for further research, and does justify the Chow-Yorke algorithm for some important mathematical programming problems. It is interesting to note that the Chow-Yorke algorithm is "finite" rather than "iterative". The zero curve γ has finite (albeit possibly large) length and is followed directly to the solution. There is no analogue of chattering, zig-zagging, or slow convergence near the solution. In fact, the concept of rate of convergence is not as relevant as the length of γ [20].

Computational results.

All the results were obtained on an IBM 370/158 using the code in [22], converted to double precision, and compiled with IBM's Fortran H Extended compiler. Unless otherwise mentioned, the answers were computed accurate to at least eight decimal places, and the starting point $a = 0$.

Problem 1. This is a linear complementarity problem

$$w - Mz = q, \quad w \geq 0, \quad z \geq 0, \quad wz = 0$$

where M is an $n \times n$ matrix and q is an n-vector given by

$$M = \begin{pmatrix} 1 & 2 & 2 & \cdots & 2 \\ 0 & 1 & 2 & \cdots & 2 \\ 0 & 0 & 1 & \cdots & 2 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}, \quad q = \begin{pmatrix} -1 \\ -1 \\ -1 \\ \vdots \\ -1 \end{pmatrix}.$$

M is copositive plus, positive semidefinite, and a P-matrix, all "nice" properties for the linear complementarity problem [13]. K. G. Murty [13] discovered this example and showed that a popular complementary pivoting method requires 2^n steps to solve it. Table 1 gives the results for the

Chow-Yorke algorithm.

Problem 2. This is a linear complementarity problem

$$w-Fz=q,\ w\geq 0,\ z\geq 0,\ wz=0$$

where F is an $n \times n$ matrix and q is an n-vector given by

$$F=M^tM=\begin{pmatrix} 1 & 2 & 2 & \cdots & 2 \\ 2 & 5 & 6 & \cdots & 6 \\ 2 & 6 & 9 & \cdots & 10 \\ \vdots & \vdots & \vdots & & \vdots \\ 2 & 6 & 10 & \cdots & 4(n-1)+1 \end{pmatrix}, \quad q = \begin{pmatrix} -1 \\ -1 \\ -1 \\ \vdots \\ -1 \end{pmatrix},$$

M is the matrix in Problem 1. F is strictly copositive, a P-matrix, symmetric, and positive definite. Y. Fathi, following Murty's work, has shown that a complementary pivoting method requires 2^n steps to solve this problem [7]. F is extremely ill-conditioned, and pivoting methods suffer a severe loss of accuracy for large n. Results for the homotopy method are in Table 2.

Problem 3. This is another linear complementarity problem with the same matrix F as in Problem 2, and

$$q=\frac{1}{2^{n+1}}\begin{pmatrix} -2^n \\ -2^n - 2^{n-1} \\ -2^n - 2^{n-1} - 2^{n-2} \\ \vdots \\ -2^n - 2^{n-1} - \cdots - 2 \end{pmatrix}.$$

This example was suggested by Murty and Fathi [7] because it gives some complementary pivoting methods difficulty. For $n > 56$, the floating point hardware on the IBM 370 cannot resolve the components of q, and thus Table 3 stops at $n = 50$. Due to the nature of $K(z)$ (recall that the complementarity problem is equivalent to $K(z)=0$) and the ill-conditioning of F, the computed answers were not accurate to eight places, although $\|K(z)\|_\infty < 10^{-8}$. Table 3 gives the results.

Problem 4. This is an unconstrained minimization problem

$$\min f(x)=100(x_2-x_1^2)^2 + (1-x_1)^2 + 90(x_4-x_3^2)^2 + (1-x_3)^2 + 10.1[(x_2-1)^2 + (x_4-1)^2] + 19.8(x_2-1)(x_4-1)$$

known as Colville's Problem Number 4, and is the first problem in Saigal [15]. This problem is fairly difficult for some minimization techniques, but was easily handled by Saigal's accelerated simplicial homotopy method [15], and the Chow-Yorke algorithm, which required 99 Jacobian evaluations, .54 seconds of CPU time, with an arc length of 2.765 . f is not convex, but using the technique in [23] it is possible to prove that the Chow-Yorke algorithm must converge to a zero of ∇f.

Problem 5. This is a 20-dimensional fixed point problem of the form

$$f_i(x)=a_i x_{i_1} x_{i_2} x_{i_3} + b_i$$

considered by Saigal [15] and Kellogg, Li, and Yorke [9]. The data

a_i, b_i, i_1, i_2, i_3 is from Saigal [15]. This problem required 46 Jacobian evaluations, and 4.8 seconds of CPU time with an arc length of 1.627.

Problem 6. This is an 80-dimensional fixed point problem of the form

$$f_i(x) = a_i x_{i_1} x_{i_2} x_{i_3} + b_i \quad ,$$

where the data a_i, b_i, i_1, i_2, i_3 is from Saigal [15]. This problem was also considered by Kellogg, Li, and Yorke [9]. The Chow-Yorke algorithm used 201 seconds of CPU time, with 52 Jacobian evaluations and arc length 2.605 .

Problem 7. The unconstrained optimization problem

$$\max f(x) = \frac{\Phi}{K_1+\Phi} \; \frac{K_2}{K_2+\theta} \quad , \quad \text{where } \Phi(x) = \sum_{i=1}^{30} a_i x_i \; ,$$

$$\theta(x) = \sum_{i=1}^{30} b_i x_i^2 \; ,$$

is from [1], and was the third problem considered in Saigal [15]. Using the raw data from [15], the zero curve γ had arc length 37.2 starting from a=(.5, ..., .5), and required 598 Jacobian evaluations and 172 seconds of CPU time. Obviously the overall efficiency of the Chow-Yorke algorithm depends on the arc length of γ , and the range of the components of the solution has some bearing on the length of γ . Components of the solution here vary from .3 to 18, resulting in the relatively long arc length. Scaling x by a factor of 10 and starting at a=0 reduces the arc length to 4.93, with 431 Jacobian evaluations in 125 seconds.

Problem 8. This is an unconstrained minimization problem:

$$\min \theta(x)=x^t[D+v(x)v(x)^t]x+q^tx$$

where D is a positive definite diagonal matrix and $v_k(x)=\sin x_k$. $\theta(x)$ is a sinusoidal oscillation superimposed on a convex quadratic function, is not convex, has many local minima and a global minimum. For each dimension n=10,20,...,100, ten different runs were made with the diagonal elements of D and the components of q chosen randonly in (0,1) and (-1,1) respectively. The results are shown in Table 4. In many cases the Chow-Yorke algorithm converged to a local minimum.

Problem 9. min $\theta(x)$ such that $x\geq 0$, where $\theta(x)$ is the same as in Problem 8. D and q were chosen randomly as in Problem 8, and ten runs were made for each dimension n=10, 20, ..., 100. Compared to the unconstrained problem, the execution time roughly doubled. The number of Jacobian evaluations required are shown in Table 5. Again in some cases only a local minimum was found.

References

[1] Albern, Ruston, and Torii, The size of rod signals, J. Physiology, 206(1970) 193-208.

[2] P. Boggs, The solution of nonlinear systems of equations by A-stable integration techniques, SIAM J. Numer. Anal., 8(1971) 767-785.

[3] S.N. Chow, J. Mallet-Paret, and J.A. Yorke, Finding zeros of maps: homotopy methods that are constructive with probability one, Math. Comp., 32(1978) 887-899.

[4] G. Dahlquist, A. Bjorck, and N. Anderson, Numerical Methods, Prentice-Hall, Englewood Cliffs, New Jersey, 1974.

[5] B.C. Eaves, Homotopies for the computation of fixed points, Math. Programming, 3 (1972), pp. 1-22.

[6] B.C. Eaves and R. Saigal, Homotopies for computation of fixed points on unbounded regions, Math. Programming, 3 (1972), pp. 225-237.

[7] Y. Fathi, Private communication, August, 1978.

[8] F.J. Gould, Recent and past developments in the simplicial approximation approach to solving nonlinear equations - a subjective view, presented at International Symposium on Extremal Methods and Systems Analysis, U. of Texas, Austin, TX, September 1977.

[9] R.B. Kellogg, T.Y. Li, and J. Yorke, A constructive proof of the Brouwer fixed-point theorem and computational results, SIAM J. Numer. Anal., 13 (1976), pp. 473-483.

[10] O.L. Mangasarian, Equivalence of the complementarity problem to a system of nonlinear equations, SIAM J. Appl. Math., 31 (1976) 89-92.

[11] O.L. Mangasarian, Nonlinear Programming, McGraw-Hill, New York, 1969.

[12] O. Merrill, Applications and extensions of an algorithm to compute fixed points of upper semicontinuous mappings, Doctoral thesis, I.O.E. Dept., University of Michigan, Ann Arbor, Michigan, 1972.

[13] K.G. Murty, Computational complexity of complementary pivot methods, Math. Programming Study 7 (1978) 61-73.

[14] J.M. Ortega and W.C. Rheinboldt, Iterative solution of nonlinear equations in several variables, Academic Press, New York, 1970.

[15] R. Saigal, On the convergence rate of algorithms for solving equations that are based on methods of complementary pivoting, Math. Operations Res., 2(1977) 108-124.

[16] H. Scarf, The approximation of fixed points of a continuous mapping, SIAM J. Appl. Math., 15(1967) 1328-1343.

[17] C.Y. Wang and L.T. Watson, Squeezing of a viscous fluid between elliptic plates, Appl. Sci. Res., to appear.

[18] L.T. Watson, T.Y. Li, and C.Y. Wang, Fluid dynamics of the elliptic porous slider, J. Appl. Mech., 45 (1978) 435-436.

[19] L.T. Watson and W.H. Yang, A homotopy method for uniform energy density structural design, SIAM J. Math. Anal., submitted.

[20] L.T. Watson, A globally convergent algorithm for computing fixed points of C^2 maps, Appl. Math. Comput., to appear.

[21] L.T. Watson, Solving the nonlinear complementarity problem by a homotopy method, SIAM J. Control Optimization, to appear.

[22] L.T. Watson and D. Fenner, Chow-Yorke algorithm for fixed points or zeros of C^2 maps, ACM Trans. Math. Software, to appear.

[23] L.T. Watson, An algorithm that is globally convergent with probability one for a class of nonlinear two-point boundary value problems, SIAM J. Numer. Anal., to appear.

n	number of Jacobian evaluations	CPU time (sec)	arc length
10	102	2.6	1.4939
20	92	9.8	1.4997
30	111	31.4	1.5023
40	112	65.5	1.5039
50	114	122.5	1.5049
60	104	186.2	1.5056
70	104	283.8	1.5062
80	104	413.1	1.5067
90	124	686.7	1.5072
100	126	937.7	1.5075

Table 1.

n	number of Jacobian evaluations	CPU time (sec)	arc length
10	106	2.8	1.4583
20	111	12.1	1.4576
30	112	32.4	1.4576
40	112	66.2	1.4575
50	114	125.0	1.4577
60	112	203.6	1.4578
70	112	305.7	1.4574
80	113	448.8	1.4579
90	113	625.4	1.4579
100	117	871.7	1.4579

Table 2.

n	number of Jacobian evaluations	CPU time (sec)	arc length	$\|\|error\|\|_\infty$
10	276	7.2	1.1855	1.02E-5
20	478	53.1	1.1857	1.86E-4
30	457	131.7	1.1856	1.96E-4
40	1100	673.8	1.1850	6.68E-5
50	850	935.0	1.1849	3.66E-5

Table 3.

run \ n	10	20	30	40	50	60	70	80	90	100
1	59	59	61	55	65	59	57	59	69	69
2	61	63	67	65	57	63	55	69	67	69
3	57	57	61	63	63	65	61	67	63	67
4	61	63	59	59	67	69	65	67	67	69
5	59	52	61	59	65	63	67	67	67	69
6	57	59	65	57	59	65	67	63	69	69
7	52	44	61	63	65	61	67	65	67	65
8	57	63	65	59	55	65	59	67	69	65
9	59	63	57	57	51	65	65	67	67	63
10	55	57	55	57	67	63	67	69	63	57

Table 4. Number of Jacobian evaluations.

run \ n	10	20	30	40	50	60	70	80	90	100
1	86	88	83	81	77	83	76	95	89	73
2	97	108	105	104	104	103	112	125	116	115
3	64	67	86	91	99	113	98	92	107	119
4	68	59	60	91	71	82	82	80	73	86
5	162	164	165	160	133	131	135	147	136	145
6	75	101	106	103	99	101	96	97	104	117
7	106	106	111	114	112	115	119	112	113	130
8	79	65	104	105	102	113	101	138	138	149
9	77	112	120	119	109	112	107	114	93	111
10	82	99	152	101	106	136	136	134	137	152

Table 5. Number of Jacobian evaluations.