Computers and Chemical Engineering 35 (2011) 403–411

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

A new homotopy for seeking all real roots of a nonlinear equation

Saeed Khaleghi Rahimian[a], Farhang Jalali[b], J.D. Seader[c,*], R.E. White[a]

[a] Department of Chemical Engineering, University of South Carolina, Columbia, SC 29208, USA
[b] Chemical Engineering Department, University of Tehran, Tehran, Iran
[c] Department of Chemical Engineering, University of Utah, Salt Lake City, UT 84112, USA

A R T I C L E   I N F O

Article history:
Received 15 March 2010
Accepted 5 April 2010
Available online 10 April 2010

Keywords:
Nonlinear equation
Homotopy
Continuation
All real roots
Bifurcation
Path tracking

A B S T R A C T

A new continuation method, which applies a new homotopy that is a combination of the fixed-point and Newton homotopies (FPN), is developed for seeking all real solutions to a nonlinear equation, written as $f(x) = 0$, without having to specify a bounded interval. First, the equation to be solved is multiplied by $(x - x^0)$, where x^0 is the starting value, which is set to zero unless the function does not exist at x^0, in which case x^0 becomes a tracking initiation point that can be set arbitrarily to any value where the function does exist. Next, the new function, $(x - x^0)f(x) = 0$, is incorporated into the FPN homotopy. The initial step establishes a single bifurcation point from which all real roots can be found. The second step ensures a relatively simple continuation path that consists of just two branches that stem from the bifurcation point and prevents the formation of any isola. By tracking the two branches of the homotopy path, all real roots are located. Path tracking is carried out with MATLAB, using the continuation toolbox of CL_MATCONT, developed by Dhooge et al. (2006), based on the work of Dhooge, Govaerts, and Kuznetsov (2003), which applies Moore–Penrose predictor-corrector continuation to track the path, using convergence-dependent step-size control to negotiate turning points and other sharp changes in path curvature. This new method has been applied, without failure, to numerous nonlinear equations, including those with transcendental functions. As with other continuation methods, $f(x)$must have twice-continuous derivatives.

1. Introduction

Mathematical models of steady-state chemical processes consist of systems of algebraic and/or transcendental equations. When all equations in the system are linear and independent of each other, only a single solution exists, which is readily obtained by robust numerical matrix methods. However, more commonly, the system of equations contains some nonlinear equations. In that case, multiple solutions may exist, where the unknowns may be real positive, real negative, and/or complex–conjugate pairs. When the unknowns are absolute temperature, absolute pressure, flow rate, and/or composition variables, only solution sets consisting of real positive values are of interest. When the composition variables are mole, mass, or volume fractions, only real positive values in the region bounded by 0 and 1 are of interest.

Occurrences of multiple solutions of interest in chemical engineering problems have been known for decades. These include:

1. Cubic equations of state, such as the Redlich–Kwong equation, where, depending on the temperature, pressure, and mixture composition, there exists three real-positive roots or one real positive root and one pair of complex–conjugate roots.
2. Fluid flow through a converging–diverging nozzle, where, for a sufficiently large flow rate, one real-positive solution is for subsonic conditions in the converging section of the nozzle and another real-positive solution is for supersonic conditions in the diverging section of the nozzle.
3. Underwood equations for component distribution from the feed to the distillate and bottoms at minimum reflux and infinite stages, where one real-positive solution exists between each pair of ordered relative volatilities.
4. Liquid–liquid equilibrium compositions when using the NRTL or UNIQUAC activity coefficient equations, where more than one set of real-positive mole fractions between 0 and 1 exist.
5. Adiabatic CSTR reactor with a single, homogeneous, exothermic reaction, where three real-positive conversions may exist.
6. Effectiveness factor for a highly exothermic reaction in a porous catalyst, where three real-positive values of the factor may exist.

For each of these classic cases, robust methods have been devised to compute the multiple solutions. For example, the well known analytical solution to the three roots of a cubic equation can be applied to the cubic equation of state.

Prior to the late 1970s, the possible existence of multiple solutions in component-separation operations, e.g. distillation, was

* Corresponding author.
E-mail address: J.Seader@utah.edu (J.D. Seader).

0098-1354/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compchemeng.2010.04.007

not imagined, despite the nonlinear nature of the equations used to model the operations. However, starting with a discovery by Magnussen, Michelsen, and Fredenslund (1979) for heterogeneous azeotropic distillation, multiple solutions have been found for a wide variety of separation operations, including: isothermal flash of a multicomponent system, binary distillation, multicomponent distillation, interlinked systems of distillation columns, homogeneous azeotropic distillation, heterogeneous azeotropic distillation, reactive distillation, temperature-swing adsorption, pressure-swing adsorption, membrane separation, and isothermal, continuous MSMPR crystallization.

When considering the possibility of multiple solutions, it is worthwhile to recognize the three types of multiplicity, as discussed by Gani and Jorgensen (1994):

1. Output multiplicity, where all input variables are specified and two or more sets of output variables are found.
2. Input multiplicity, where one or more output variables are specified and two or more sets of unspecified input variables are found.
3. Internal-state multiplicity, where two or more sets of internal conditions or profiles are found for the same values of the input and output variables.

All three types of multiplicity have been found for component-separation operations.

Usually, multiplicity occurs only for certain sets of specifications and only over restricted ranges of certain parameters. Often, sets of multiple solutions contain both stable and unstable solutions. One or more of the stable solutions is clearly superior for practical application and one or more of the other stable solutions is trivial or very undesirable for practical application. For design purposes, it is important to be aware of the possibility of multiplicity and to discover all multiple solutions within the practical domain of operating variables.

Numerical methods that seek to find multiple solutions are:

1. Newton's method with deflation, e.g. Press, Flannery, Teukolsky, and Vetterling (1992).
2. Parallel-path homotopy-continuation methods for systems of polynomial equations, which use Bezout's theorem for determining the maximum number of multiple solutions and finds all solutions, real and positive, e.g. Morgan (1987).
3. Global homotopy-continuation methods, which attempt to find all sets of real solutions, e.g. references below.
4. Interval Newton method with generalized bisection, which locates all real roots within specified intervals of the unknowns, e.g. Kearfott and Novoa (1990).
5. Global terrain methods, which consist of a series of downhill, equation-solving computations and uphill predictor-corrector calculations to find all physically meaningful solutions and singular points, e.g. Lucia and Feng (2002).
6. Deterministic branch-and-bound method that transforms the system of equations into a global optimization problem, e.g. Maranas and Floudas (1995).
7. Methods using cellular exclusion tests, e.g. Georg (2003).

Method 1 is unreliable. Method 2 is very reliable and easy to use, but is restricted to polynomial equations. Methods 4–7 are reliable, but require the specification of bounded intervals.

Here, we concentrate on Method 3, which has been discussed and applied by a number of investigators, including Keller (1977, 1978), Chow, Mallet-Paret, and Yorke (1978), Garcia and Zangwill (1979), Allgower and Georg (1980), Watson (1986), Wayburn and Seader (1987), Allgower and Georg (1987), Kuno and Seader (1988), Seader et al. (1990), Sun and Seider (1995), Watson, Sosonkina, Melville, Morgan, and Walker (1997), Kuznetsov (1998), Jalali and Seader (1999), Bausa and Marquardt (2000), Gritton, Seader, and Lin (2001), Wu (2005), and Imai, Yamamura, and Inoue (2005). In this paper, only a single nonlinear algebraic and/or transcendental equation is considered, and a new homotopy is introduced and applied to the determination of all real roots. Unlike the method of Gritton, Seader, and Lin, the homotopy path of our method never consists of branches that are only connected at opposite infinities, or by branches in the complex domain. Our new homotopy only consists of one path, which is quickly tracked in the forward and backward directions, called Branches 1 and 2, from the starting point or a tracking initiation point. In a subsequent paper, our new homotopy is extended to determine all real solutions to systems of nonlinear algebraic and/or transcendental equations.

2. The new homotopy (FPN)

A widely used homotopy, $H(x, t)$, consists of a linear combination of two real functions: $f(x)$, whose zeroes are sought; and $G(x)$, a function for which a zero is known or readily selected or obtained. Both functions must be smooth with twice-differentiable derivatives. Thus,

$$H(x, t) = tf(x) + (1 - t)G(x) = 0 \tag{1}$$

where t, the homotopy parameter, allows tracking of a solution path that connects the starting point, x^0, at $t = 0$, to all x_i^*, which are all solutions of $f(x) = 0$. Using numerical continuation, the parameter t is gradually varied, starting from $t = 0$ and without being confined, leading to a series of solutions to Eq. (1). Whenever the homotopy path crosses $t = 1$, a solution to $f(x) = 0$ is found.

When attempting to determine all roots of $f(x) = 0$, the choice of $G(x)$ can be important. The three most widely cited $G(x)$ functions are the fixed-point (FP) function,

$$G(x) = (x - x^0), \tag{2}$$

the affine function, which adds a factor, A, to the FP function to improve scaling in the homotopy function,

$$G(x) = A(x - x^0), \tag{3}$$

where A is often taken as the derivative of $f(x)$ evaluated at x^0, and the Newton (N) function,

$$G(x) = f(x) - f(x^0) \tag{4}$$

When Eqs. (2) or (3) are applied to Eq. (1), the homotopy is satisfied by only one root at $t = 0$, the usual starting point. However, the path can consist of branches that are only connected at opposite infinities or by branches in the complex domain. Kuno and Seader (1988) present two methods for avoiding the infinity problem. One method involves a criterion for establishing a starting point; but the criterion cannot always be implemented. The second method involves an auxiliary function, which, however, adds an additional equation to the one to be solved. When Eq. (4) is applied, $f(x) = f(x^0)$ at $t = 0$ and multiple starting points and multiple branches may exist. Here, for reasons discussed below, we introduce a new formulation of the homotopy function of Eq. (1), which incorporates a new $G(x)$ function. At first, the new homotopy appears to be considerably more complicated than Eq. (1) with Eqs. (2), (3), or (4). However, after simplification, the new homotopy is less complex, avoids the disadvantages of the fixed-point, affine, and Newton functions, and possesses superior characteristics for tracking the homotopy path to determine all solutions to $f(x) = 0$. We refer to the new homotopy, as the FPN homotopy because it uses elements of the fixed-point and Newton functions. Most important, the FPN homotopy does not have multiple starting points and the path is not connected at opposite infinities.