# A NOTE ON THE RELATION BETWEEN THE NEWTON HOMOTOPY METHOD AND THE DAMPED NEWTON METHOD*

XUPING ZHANG[†] AND BO YU[†]

**Abstract.** The homotopy continuation method and the damped Newton method are two known methods for circumventing the drawback of local convergence of the standard Newton method. Although some relations between these two methods have already been obtained, these relations are mainly for the differential equations which determine the paths followed by the two methods, rather than the sequences generated by the algorithms. In this paper, these sequences are investigated and some further relations are explored in terms of the marching directions and the step sizes during the iteration processes. Numerical solution of a semilinear elliptic equation is included to illustrate the relations discovered.

**Key words.** Homotopy continuation, damped Newton method, domain of convergence, nonlinear algebraic equations, semilinear elliptic equations, finite element method.

**AMS subject classifications.** 65H10, 65H20.

**1. Introduction.** The primary goal of this paper is to present recently observed relations between the homotopy continuation and the damped Newton methods for nonlinear equations $F(x) = 0$, where $F$ is a nonlinear map from a Banach space $X$ to a Banach space $Z$. Specifically, the homotopy considered in this paper is the so called Newton homotopy,

$$(1.1) \qquad H(x,t) = F(x) + (t-1)F(x^0), \quad t \in [0,1].$$

Some results on such relations were derived earlier; see for example [5] and [3]. These papers are mainly concerned with the relations between the differential equations that determine the paths followed by the algorithms. However, what mainly interests us are the relations between the two sequences generated by the predictor-corrector (PC) homotopy continuation method and the damped Newton method, since in actual implementation these two sequences generally deviate somewhat from the continuous paths. The iteration processes will be investigated and relations will be explored in terms of the marching directions and the step sizes.

The paper is organized as follows. In Section 2, the explicit Euler method for an ordinary differential equation (ODE) initial value problem (IVP) is investigated under a parameter transformation. In Section 3, the relations between the PC homotopy continuation method and the damped Newton method are explored under the same space setting. The numerical performance is illustrated in Section 4 and concluding remarks are presented in Section 5.

**2. The explicit Euler method under parameter transformation.** For clarity, in the following we denote the sequence generated by the homotopy continuation method as $(x^k, t_k)$ and the sequence generated by the damped Newton method as $(y^k, \lambda_k)$; $(\bar{y}^k, t_k)$ denotes the same sequence after the parameter transformation. As it is known, the damped Newton method

$$y^{k+1} = y^k - \Delta\lambda_k F'(y^k)^{-1} F(y^k), \quad y^0 = x^0,$$

---

can be considered an explicit Euler method for the following ODE IVP,

(2.1)
$$\begin{cases} \dfrac{dy}{d\lambda} = -F'(y)^{-1}F(y), & \lambda \in [0, +\infty), \\ y(0) = x^0. \end{cases}$$

In conventional terms, the solution $y(\lambda)$ of (2.1) is referred to as a phase trajectory, and $(y(\lambda), \lambda)$ as an integral curve. From the ODE in (2.1), it holds that

$$F'(y)\frac{dy}{d\lambda} + F(y) = 0,$$

which amounts to

$$\frac{d}{d\lambda}\left(e^\lambda F(y(\lambda))\right) \equiv 0.$$

Thus, in the PC homotopy continuation framework, this integral curve $(y(\lambda), \lambda)$ is the path determined by the homotopy

(2.2)      $$H(y, \lambda) = F(y) - e^{-\lambda}F(x^0) = 0, \quad (y, \lambda) \in X \times [0, +\infty).$$

Since, conventionally, the domain of the imbedding parameter in the homotopy continuation method is $[0, 1]$, the parameter transformation

$$t = \varphi(\lambda) = 1 - e^{-\lambda},$$

is needed. By this transformation, the homotopy (2.2) is changed to

$$H(\bar{y}, t) = F(\bar{y}) + (t - 1)F(x^0) = 0, \quad (\bar{y}, t) \in X \times [0, 1],$$

which is essentially the same as the homotopy (1.1). It is therefore necessary to consider the explicit Euler method under parameter transformation for an ODE IVP.

Consider the ODE IVP

(2.3)
$$\begin{cases} \dfrac{du}{d\lambda} = f(\lambda, u), \\ u(0) = u^0. \end{cases}$$

Under the transformation

$$\lambda = \psi(t), \quad 0 = \psi(0),$$

the ODE IVP (2.3) is transformed to

(2.4)
$$\begin{cases} \dfrac{d\bar{u}}{dt} = \psi'(t)\bar{f}(t, \bar{u}), \\ \bar{u}(0) = u^0. \end{cases}$$

The explicit Euler methods for (2.3) and (2.4) are respectively

(2.5)                    $$u^{k+1} = u^k + \Delta\lambda_k f(\lambda_k, u^k)$$

and

(2.6)                    $$\bar{u}^{k+1} = \bar{u}^k + \Delta t_k \psi'(t_k)\bar{f}(t_k, \bar{u}^k).$$

If $\lambda_k = \psi(t_k)$ and $\lambda_{k+1} = \psi(t_{k+1})$, then in general $\Delta\lambda_k \neq \Delta t_k \psi'(t_k)$, and the sequences $\{u^k\}$ and $\{\bar{u}^k\}$ generated by (2.5) and (2.6) are different. In this sense, the explicit Euler method is not invariant under a parameter transformation. In order for $\{u^k\}$ and $\{\bar{u}^k\}$ to be the same, $\Delta t_k$ should be chosen such that $\Delta\lambda_k = \Delta t_k \psi'(t_k)$, in which case $\lambda_k$ is different from $\psi(t_k)$.

**3. The relations.** In order to compare the damped Newton method with the PC homotopy continuation method, it is transformed to the underlying space $X \times [0, 1]$. Under the parameter transformation

$$\lambda = -\ln(1 - t),$$

the ODE IVP (2.1) is changed to

$$(3.1) \qquad \begin{cases} \dfrac{d\bar{y}}{dt} = -\dfrac{1}{1-t}F'(\bar{y})^{-1}F(\bar{y}), & t \in [0, 1), \\ \bar{y}(0) = x^0. \end{cases}$$

The damped Newton method in $X \times [0, \infty)$,

$$(3.2) \qquad y^{k+1} = y^k - \Delta\lambda_k F'(y^k)F(y^k), \quad y^0 = x^0,$$

interpreted as the explicit Euler method for (2.1), is changed to its counterpart in $X \times [0, 1]$,

$$(3.3) \qquad \bar{y}^{k+1} = \bar{y}^k - \frac{\Delta t_k}{1 - t_k}F'(\bar{y}^k)^{-1}F(\bar{y}^k), \quad \bar{y}^0 = x^0,$$

which is the explicit Euler method for (3.1). In order for $\bar{y}^k = y^k$ in the phase space $X$, $\Delta t_k$ should be chosen such that $\frac{\Delta t_k}{1-t_k} = \Delta\lambda_k$. This section is mainly focused on the relations between the sequence $\{\bar{y}^k\}$, generated by (3.3), and the sequence $\{x^k\}$ generated by the PC homotopy continuation method for (1.1).

Assume for simplicity that the path determined by (1.1) is $(x(t), t)$. Then $x(t)$ satisfies the following ODE IVP

$$(3.4) \qquad \begin{cases} \dfrac{dx}{dt} = -F'(x)^{-1}F(x^0), & t \in [0, 1], \\ x(0) = x^0. \end{cases}$$

In a continuous framework, the ODE IVPs (3.1) and (3.4) are equivalent, as the phase trajectories determined by them are the same. Since $F'(x)\frac{dx}{dt} + F(x^0) = 0$ is used to compute the tangent vectors in the PC method, the explicit Euler method for (3.4) is equivalent to the homotopy continuation method for the homotopy (1.1), where only the predictor step is performed. Such a continuation method will be called a *no-corrector* homotopy continuation, for convenience. Therefore, it may seem that the damped Newton method (3.2) is nothing more than the *no-corrector* homotopy continuation method. However, this is not true. When the explicit Euler method is employed, the ODE IVPs (3.1) and (3.4) are not equivalent anymore, since the vector fields of the two ODEs are different, and so the sequences generated by them will not be the same. This will be discussed in Section 3.1.

**3.1. Relation of marching directions.** Recall the assumption that the path determined by the homotopy (1.1) is $(x(t), t)$, and the tangent vectors in the PC method are computed by $F'(x)\frac{dx}{dt} + F(x^0) = 0$.
*First iteration.* For the PC method the first tangent $(d^0, \tau_0)$ at $(x^0, 0)$ is

$$(3.5) \qquad d^0 = -F'(x^0)^{-1}F(x^0), \quad \tau_0 = 1.$$

After one PC iteration, $(x^1, t_1)$ is generated. We can see that $d^0$ in (3.5) is just the Newton direction. If the same step sizes are chosen, and the corrector step is not performed, then the same iterate $x^1 = \bar{y}^1$ is obtained.

*Second iteration.* The tangent $(d^1, \tau_1)$ at $(x^1, t_1)$ can be computed as follows. If $(x^1, t_1)$ lies on the path $H^{-1}(0)$ exactly, then

$$\begin{cases} d^1 = -F'(x^1)^{-1}F(x^0) = -\dfrac{1}{1-t^1}F'(x^1)^{-1}F(x^1), \\ \tau_1 = 1. \end{cases}$$

If $(x^1, t_1)$ does not lie on the path $H^{-1}(0)$ exactly and $H(x^1, t_1) = c_1 \neq 0$, then $d^1$ is instead

(3.6)
$$\begin{aligned} d^1 &= -F'(x^1)^{-1}F(x^0) = -\frac{1}{1-t^1}F'(x^1)^{-1}\left(F(x^1) - c_1\right) \\ &= -\frac{1}{1-t^1}F'(x^1)^{-1}F(x^1) + \frac{1}{1-t^1}F'(x^1)^{-1}c_1. \end{aligned}$$

In fact, this is the tangent on the path defined by $H(x, t) = c_1 \neq 0$. From (3.6) it is obvious that, even if $x^1 = \bar{y}^1$, the tangent component $d^1$ in the PC method is no longer the Newton direction of $F$ at $x^1$.

By iterating, it can be seen that the damped Newton sequence $\{\bar{y}^k\}$ is not generally the same as the sequence $\{x^k\}$ generated by the *no-corrector* homotopy continuation method. Using the terminology of dynamical system, the damped Newton method is marching along the vector field $-1/(1-t)F'(x)^{-1}F(x)$, while the homotopy continuation method is marching along the vector field $-F'(x)^{-1}F(x^0)$. The integral curves for $-1/(1-t)F'(x)^{-1}F(x)$ are the paths implicitly determined by

(3.7)                           $\bar{H}(x, t) = F(x) + (t-1)F(c_1) = 0,$

while those for $-F'(x)^{-1}F(x^0)$ are

(3.8)                           $H(x, t) = F(x) + (t-1)F(x^0) = c_2,$

where $c_1$ and $c_2$ are generic constant vectors. These two families of paths are different, and the common path is determined by $F(x) + (t-1)F(x^0) = 0$, i.e., $c_1 = x^0$ and $c_2 = 0$.

The paths determined by (3.7) and (3.8) are illustrated in Figure 3.1, where the solid lines are paths corresponding to (3.7), the dotted lines corresponding to (3.8) and the thick line to the common path. As illustrated in Figure 3.1, the directions along which the damped Newton method marches are the tangents to the solid lines, while the directions of prediction in the homotopy continuation method are the tangents to the dotted lines. The Newton directions are good directions, since even though the damped Newton iterates may deviate from the path that the method is supposed to follow, they will lie in a curvy cone formed by the paths of $F(x) + (t-1)F(c_1) = 0$, and will bend back to the target point at the end. Thus, even though there may be some singular point on the path of $F(x) + (t-1)F(x^0) = 0$, the damped Newton iterates may steer away from it. Also, there is no need of a corrector step in the damped Newton method, therefore a lot of computational cost can be saved. However, for the PC homotopy continuation method a corrector step is of essential importance. If a correction is performed, the iterates of the PC method will stick to the path determined by the homotopy $H(x, t) = F(x) + (t-1)F(x^0) = 0$, otherwise the iterates will always deviate from the path and never bend back to the target point.

**3.2. Relation of step sizes.** First of all, the notion of step size needs to be clarified, since the definitions of step size are different in the PC method and the damped Newton method. In the PC method, when the homotopy path is parameterized with respect to arc length $s$, the step size $h$ in the predictor step is understood as

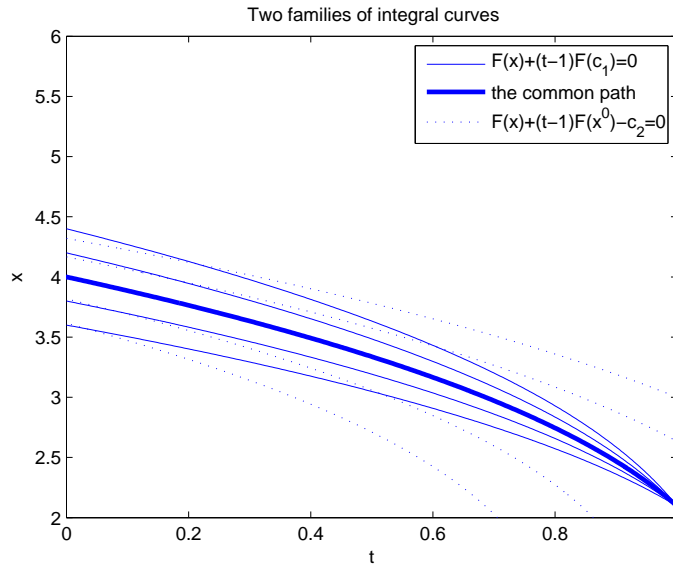$$(\tilde{x}^{k+1}, \tilde{t}_{k+1}) = (x^k, t_k) + h(d^k, \tau_k),$$

FIGURE 3.1. *Two families of paths followed by the PC and damped Newton methods.*

where $(d^k, \tau_k)$ is the unit tangent. When the path can be parameterized with respect to $t$, the step size $h$ in the predictor step is usually referred to as $\Delta t_k$:

$$(3.9) \qquad (\tilde{x}^{k+1}, \tilde{t}_{k+1}) = (x^k, t_k) + \Delta t_k \left( \frac{dx}{dt}(t_k), 1 \right).$$

In the damped Newton method the step size $\Delta \lambda_k$ instead occurs as

$$y^{k+1} = y^k - \Delta \lambda_k F'(y^k)^{-1} F(y^k).$$

In this paper, for the PC method tracing the homotopy path $(x(t), t)$, we assume the definition (3.9). As mentioned in Section 2, in order that $\bar{y}^k = y^k$, $\Delta t_k$ in (3.3) should be chosen such that

$$(3.10) \qquad \frac{\Delta t_k}{1 - t_k} = \Delta \lambda_k.$$

With the relation (3.10) and the condition $0 < \Delta \lambda_k < 1$, it is easily derived by induction that

$$\Delta t_k = \Delta \lambda_k \prod_{i=0}^{k-1} (1 - \Delta \lambda_i).$$

If $\Delta \lambda_k \equiv \Delta \lambda$, then

$$\Delta t_k = (1 - \Delta \lambda)^k \Delta \lambda.$$

On the other hand, if $\Delta t_k \equiv \Delta t$, then setting $K = 1/\Delta t$ it can be shown that

$$\Delta \lambda_k = \frac{\Delta t}{1 - k\Delta t} = \frac{1}{K - k}.$$

Thus, the damped Newton method with step size $\Delta \lambda_k \equiv \Delta \lambda$ approximately corresponds to the PC method with step size $\Delta t_k = (1 - \Delta \lambda)^k \Delta \lambda$, while the PC method with step size $\Delta t_k \equiv \Delta t$ approximately corresponds to the damped Newton method with step size $\Delta \lambda_k = 1/(K - k)$.

**4. Numerical illustrations.** The test problem we consider is the Bratu problem

$$(4.1) \qquad \begin{cases} -\Delta u = 2.5e^u, & x \in \Omega = (0,1) \times (0,1), \\ u = 0, & x \in \partial\Omega. \end{cases}$$

The initial guess

$$u_0 = 2.4 \sin(\pi x)\sin(\pi y) + 1.985 \cdot 2.4 \sin(10\pi x)\sin(10\pi y)$$

is chosen so far away from the solution that the standard Newton method does not converge. The initial guess and the solution are plotted in Figure 4.1. The differential equation is discretized by the $P1$ finite element method on a triangular mesh with node number $N = 1089$ and free node number $n = 961$. Let $F(U) = 0$ be the discretized equations of (4.1) and $H(U,t) = F(U) + (t-1)F(U^0)$ be the homotopy, where $U^0$ corresponds to the initial guess function $u_0$. The path which the homotopy continuation method and the damped Newton method are supposed to follow is determined by $H(U,t) = 0$. For details on the finite element method, see [2].
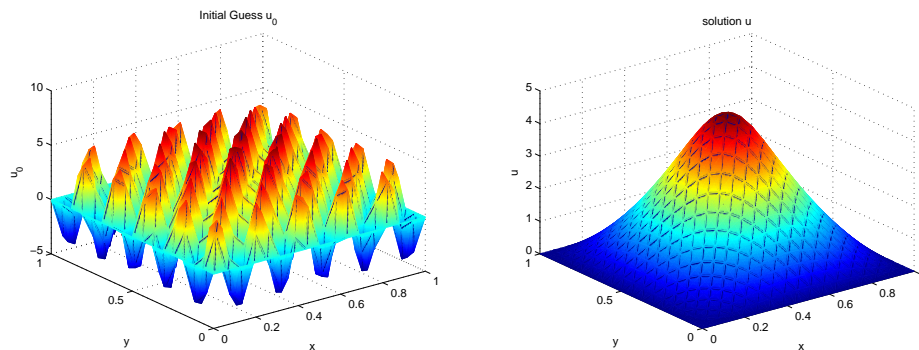


FIGURE 4.1. *Initial guess $u_0$ and solution u.*

The damped Newton method is applied with step sizes $\Delta\lambda_k \equiv 0.01$ and $\Delta\lambda_k = 1/(K - k)$. Here $K = 50$. The norm of the residual versus the number of iterations is plotted in Figure 4.2. It can be seen that with $\Delta\lambda_k \equiv 0.01$ the convergence rate of the damped Newton method becomes very slow at the last stage of the iteration process. This phenomenon can be explained, from the viewpoint of the homotopy continuation method, by noting that $\Delta\lambda_k \equiv 0.01$ in the damped Newton method corresponds approximately to $\Delta t_k = (1 - \Delta\lambda)^k\Delta\lambda \to 0$ in the homotopy continuation method. On the other hand, with $\Delta\lambda_k = 1/(K - k)$ the convergence rate is almost uniform, since $\Delta t_k$ is approximately $1/K$ in the homotopy continuation method.

To see the difference in marching directions, the performance of the *no-corrector* homotopy continuation method and the damped Newton method is shown in Figure 4.3, in terms of the angle between tangents versus the number of iterations.

As mentioned in Section 3.1, the iterates of the homotopy continuation method will follow the path $H(U,t) = 0$. There are two simple bifurcation points on this path detected and jumped over by the homotopy continuation method, in which the "Jumping Over A Bifurcation Point" algorithm of [1] is implemented. A simple bifurcation point $(\hat{x}, \hat{t})$ is characterized by $\mathrm{rank}(H_U(\hat{x}, \hat{t})) = n - 1$ and $\mathrm{rank}([H_U(\hat{x}, \hat{t}), H_t(\hat{x}, \hat{t})]) = n - 1$. The essence of the algorithm is as follows. The direction of the tangent $(d, \tau)$ of the homotopy path $H(U,t) = 0$
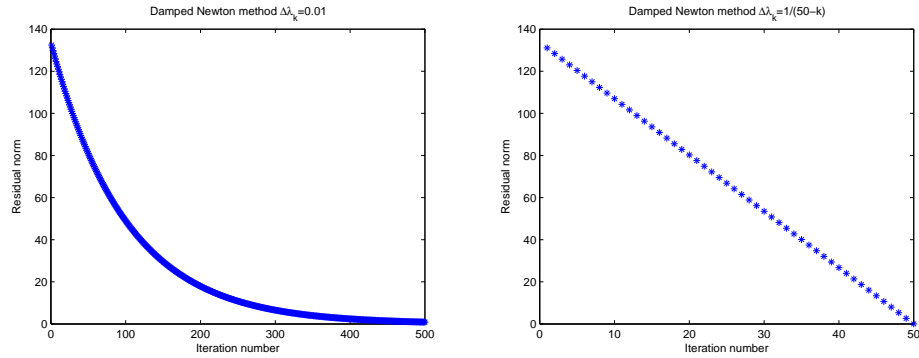
FIGURE 4.2. *Residual norms by damped Newton method, left:* $\Delta\lambda_k = 0.01$; *right:* $\Delta\lambda_k = 1/(50 - k)$.
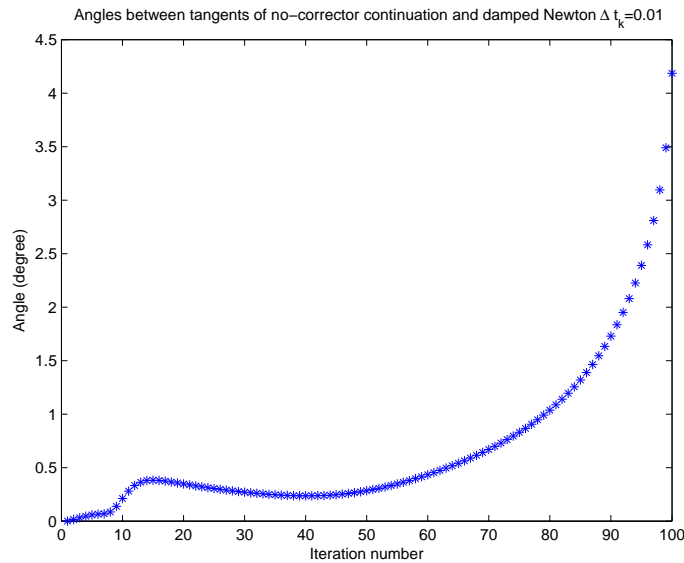
FIGURE 4.3. *Angles of tangents of* no-corrector *homotopy continuation method and damped Newton method.*

is determined by the following orientation at the starting point $(U^0, 0)$,

$$(4.2) \qquad \qquad \text{sign}\left(\det\begin{bmatrix} H_U & H_t \\ d^T & \tau \end{bmatrix}\right).$$

During the iteration process, if at two successive iterates $(x^k, t^k)$ and $(x^{k+1}, t^{k+1})$ the angle of their tangents $(d^k, \tau^k)$ and $(d^{k+1}, \tau^{k+1})$ is greater than $90°$, then there must be a bifurcation point between the two iterates along the path, and the orientation (4.2) is reversed in order to march ahead to reach $t = 1$. For example, one of such bifurcation point is roughly at $t = 0.137274$, where the angle of the tangents is nearly $180°$ and (numerically) $\text{rank}(H_U) = \text{rank}([H_U, H_t]) = n - 1 = 960$, which characterizes a simple bifurcation point. For more details on detection of bifurcation and continuation past bifurcation, see [4] and [1]. On the other hand, the iterates of the damped Newton method deviated from the path at its beginning, and bent back at its end. Thus, the iterates steered away from these two

bifurcation points. If turning points with respect to $t$ were encountered, the damped Newton method would break down, while the pseudo-arclength homotopy continuation method could pass through. Furthermore, thanks to the corrector, the maximum of allowed uniform step sizes which assures convergence for the homotopy continuation method is $\Delta t = 0.2$, larger than $\Delta t = 0.08$ for the damped Newton method. In these senses, the homotopy continuation method is more robust than the damped Newton method.

**5. Concluding remarks.** The Newton homotopy continuation method and the damped Newton method are closely related. After a transformation to the same space $X \times [0, 1]$, the homotopy continuation method marches along the vector field $-F'(x)^{-1}F(x^0)$, while the damped Newton method marches along $-\frac{1}{1-t}F'(x)^{-1}F(x)$. In addition, the step sizes of both methods have an approximate correspondence. The damped Newton method is not equivalent to the *no-corrector* homotopy continuation method. The Newton directions are good directions and there is no need for a corrector, so there is a saving in the computational cost. For the homotopy continuation method, the corrector is of essential importance for convergence and robustness. Relatively speaking, the homotopy continuation method is more robust, while the damped Newton method is more efficient.

REFERENCES

[1] E. L. ALLGOWER AND K. GEORG, *Introduction to Numerical Continuation Methods*, Classics in Applied Mathematics, 45, SIAM, Philadelphia, 2003.
[2] L. CHEN, *iFEM: an integrated finite element methods package in MATLAB*, Technical report, Department of Mathematics, University of California at Irvine, CA, November 2008.
[3] C. B. GARCIA AND F. J. GOULD, *Relations between several path following algorithms and local and global Newton methods*, SIAM Rev., 22 (1980), pp. 263–274.
[4] H. B. KELLER, *Lectures on Numerical Methods In Bifurcation Problems*, Springer, Berlin, 1986.
[5] G. H. MEYER, *On solving nonlinear equations with a one-parameter operator imbedding*, SIAM J. Numer. Anal., 5 (1968), pp. 739–752.