21

Newton-Like Methods

TABLE OF CONTENTS

§21.1. Introduction

The conventional Newton method (CNM) described in Chapter 20 is hindered by two major shortcomings:

High cost. The tangent stiffness matrix $\mathbf{K}^k = \mathbf{K}(\mathbf{u}^k, \lambda^k)$ has to be formed and factored at each iteration step.

Low Reliability. Convergence to the desired solution is not guaranteed unless the initial estimate is sufficiently close. The method may diverge or converge to an unwanted solution. This is quite likely in the vicinity of critical points.

Because of these shortcomings many variations of CNM, collectively called *Newton-like methods*, have been proposed and implemented with varying degree of success. Among the most important

1. Relaxed Newton Methods (RNM): reliability
2. Damped Newton Methods (DNM): for reliability
3. Modified Newton Methods (MNM): for efficiency
4. Quasi-Newton Methods (QNM): for efficiency

As can be seen by the large number of variations, no modification can be said to be uniformly superior to others. The above list covers the most important so-called *Newton-like methods*. Question arise, however, as to how far the offsprings can deviate from the parent and still be called Newton-like. Authors have different opinions in this matter. To further complicate things, combinations of these techniques are often used in advanced nonlinear solvers.

Some of these variants, notably the Relaxed Newton methods (RNM), are more easily derived by interpreting the Newton method in the context of a dynamic system. This interpretation is discussed next.

§21.2. The Newton Method as a Dynamical System

Figure 21.1 shows graphically what the goal of the Newton corrector is: to allow large incremental steps by eliminating the drift error.

The incremental phase is driven by the first order rate form $\dot{\mathbf{r}} = \mathbf{0}$ where the pseudo-time t is measured by an "increment clock." Penalize the drift error by adding a term proportional to the residual $\mathbf{r}$:

$$\dot{\mathbf{r}} + \mathbf{W}\mathbf{r} = \mathbf{0} \tag{21.1}$$

where $\mathbf{W}$ is a positive-definite *residual weighting matrix*, which for the moment is left arbitrary. This is called a *first order corrective form*, and also a *first order relaxation form*. It obviously reduces to $\dot{\mathbf{r}} = \mathbf{0}$ on an equilibrium path $\mathbf{r} = \mathbf{0}$. The job of the *penalty term* $\mathbf{W}\mathbf{r}$ is to force the solution trajectories of (21.1) to approach $\mathbf{r} = 0$ as the pseudotime clock t runs along a "corrective clock" See Figure 21.2.

Figures 21.1 and 21.2 are a bit deceptive in that they depict corrective processes for a one-DOF problem. A more typical state of affairs can be observed in Figure 21.3, which depicts trajectories of a corrective process on a constraint surface.

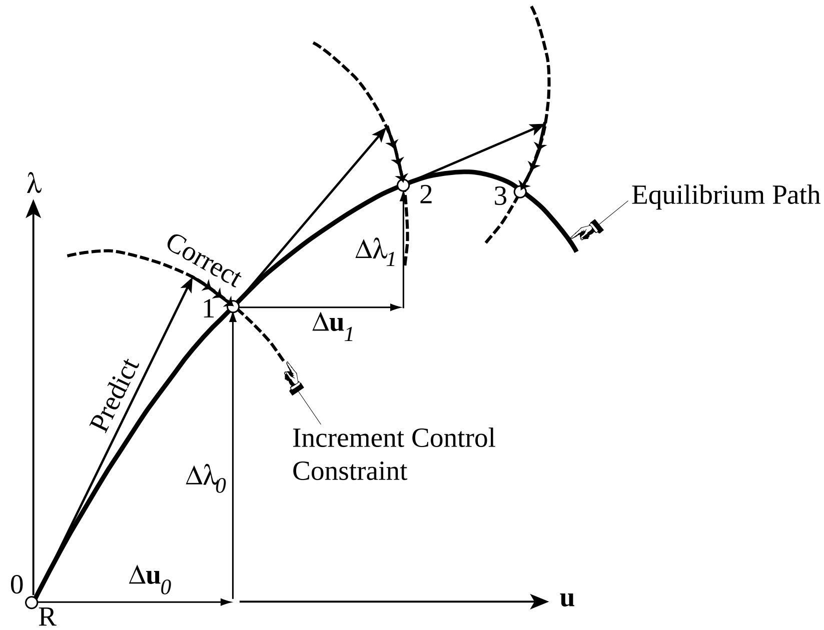

Figure 21.1. An incremental-iterative solution method.

To bring explicitly the stiffness matrix and incremental load vector into play, insert $\dot{\mathbf{r}} = \mathbf{K}\dot{\mathbf{u}} - \mathbf{q}\dot{\lambda}$ into the above and transfer $\mathbf{q}\dot{\lambda}$ to the right hand side:

$$\mathbf{K}\dot{\mathbf{u}} + \mathbf{W}\mathbf{r} = \mathbf{q}\dot{\lambda}. \tag{21.2}$$

§21.2.1. Corrective Process for Fixed λ

Suppose that we are at $\{\mathbf{u}^k, \lambda^k\}$ at which the tangent stiffness $\mathbf{K}^k$ is nonsingular. We want to move to a new state $\{\mathbf{u}^{k+1}, \lambda^k\}$ closer to $\mathbf{r} = \mathbf{0}$ while keeping $\lambda = \lambda^k$ fixed. This can be done by treating (21.2) with the Forward Euler integrator

$$\mathbf{u}^{k+1} = \mathbf{u}^k + h\dot{\mathbf{u}}^k. \tag{21.3}$$

where h is the *integration steplength*. The integrated corrector equation is

$$\mathbf{u}^{k+1} = \mathbf{u}^k - h\mathbf{F}^k\mathbf{W}^k\mathbf{r}^k, \tag{21.4}$$

where $\mathbf{F} = \mathbf{K}^{-1} = (\partial\mathbf{r}/\partial\mathbf{u})^{-1}$ is a flexibility matrix. Calling $\mathbf{d} = \mathbf{u}^{k+1} - \mathbf{u}^k$ the correction in displacements and passing $\mathbf{K}^{-1}$ to the right hand side,

$$\mathbf{K}^k\mathbf{d} = -h\mathbf{W}^k\,\mathbf{r}^k \tag{21.5}$$

If now we take

$$\mathbf{W} = \mathbf{I}, \qquad h = 1 \qquad \text{for any } k, \tag{21.6}$$

we obtain the conventional Newton method (CNM) for fixed λ, as can be easily verified. A variant called Relaxed Newton, discussed below, results by letting h be adjustable.

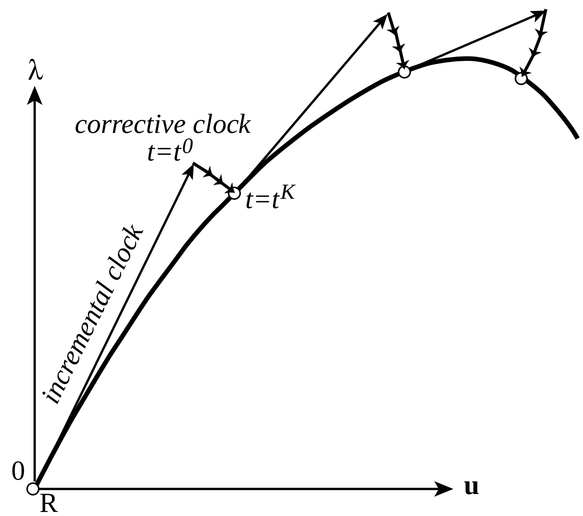

Figure 21.2. Pseudo time t running along an "incremental clock" interspersed by a "corrective clock." K is the total number of corrective iterations.

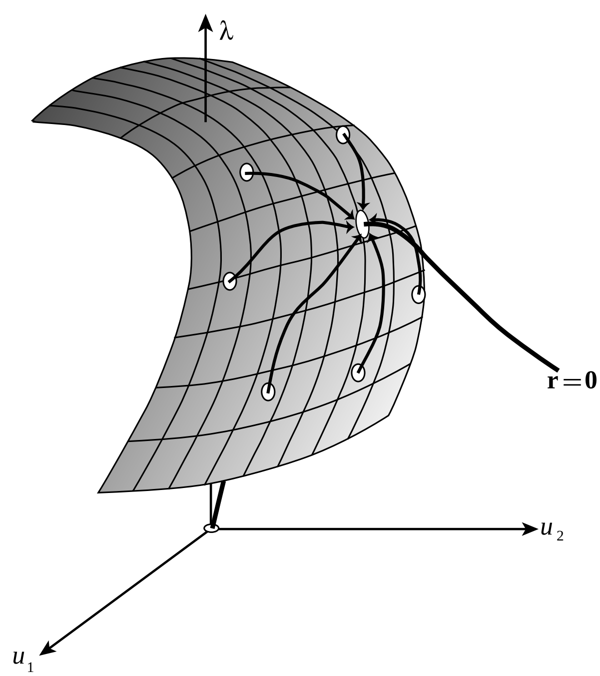

Figure 21.3. The corrective process for two degrees of freedom. The challenge is to end at a solution no matter where one starts on the constraint surface.

§21.2.2. Corrective Process for Varying λ

Suppose next that λ is to be let vary while satisfying the scalar constraint $c(\Delta\mathbf{u}, \Delta\lambda) = 0$ that controls the increment size. The corrective equation can be generalized as

$$\begin{bmatrix} \dot{\mathbf{r}} + \mathbf{W}\mathbf{r} \\ \dot{c} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 0 \end{bmatrix} \tag{21.7}$$

Inserting $\dot{\mathbf{r}} = \mathbf{K}\dot{\mathbf{u}} - \mathbf{q}\dot{\lambda}$ and $\dot{c} = \mathbf{a}^T\dot{u} + g\dot{\lambda}$, in which $\mathbf{a} = \partial c/\partial\mathbf{u}$ and $g = \partial c/\partial\lambda$, one gets

$$\begin{bmatrix} \mathbf{K}\dot{\mathbf{u}} + \mathbf{W}\mathbf{r} \\ \mathbf{a}^T\dot{\mathbf{u}} + g\dot{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{q}\dot{\lambda} \\ 0 \end{bmatrix} \tag{21.8}$$

Treat (21.8) with the Forward Euler integrator on both $\mathbf{u}$ and λ:

$$\mathbf{u}^{k+1} = \mathbf{u}^k + h\dot{\mathbf{u}}^k, \qquad \lambda^{k+1} = \lambda^k + h\dot{\lambda}^k \tag{21.9}$$

Integrating (21.8) with (21.9), followed by setting $\mathbf{W} = \mathbf{I}$ and $h = 1$, yields the conventional Newton method for general increment control treated in the previous Chapter. The verification is the matter of an exercise.

§21.3. Relaxed Newton Methods

One commonly used variant of CNM aims to increase the *reliability* but not necessarily lower the cost per iteration. This is done by deriving CNM from the dynamical process described in the previous section, and letting the steplength h be a variable. The

$$\begin{bmatrix} \mathbf{K} & -\mathbf{q} \\ \mathbf{a}^T & g \end{bmatrix} \begin{bmatrix} \mathbf{d} \\ \eta \end{bmatrix} = -\begin{bmatrix} \mathbf{r} \\ c \end{bmatrix}, \tag{21.10}$$

where supercript k is suppressed from $\mathbf{K}$, $\mathbf{q}$, etc., to reduce clutter. Solving for $\mathbf{d}$ and η as explained in the previous Chapter, one then corrects

$$\mathbf{u}^{k+1} = \mathbf{u}^k + h\mathbf{d}, \qquad \lambda^{k+1} = \lambda^k + h\eta. \tag{21.11}$$

It is understood that h may change from integration to iteration, that is, $h = h^k$.

The method (21.10)-(21.11) is called the *relaxed Newton-Raphson* method, or RNR.[1] There are three possibilities as regards h^k:

1. If $h^k < 1$, the iteration step is said to be *underrelaxed* and h^k is an *underrelaxation parameter*.
2. If $h^k > 1$, the iteration step is said to be overrelaxed and h^k is an *overrelaxation parameter*.
3. If $h^k = 1$ for all k, RNM reduces to CNM.

How is the steplength h chosen? Rules to this effect are discussed in Chapter 22.

[1] Some authors called this the *damped Newton-Raphson* method but that name is reserved here for the variant discussed in the next section.

§21.4. Damped Newton Methods

The Relaxed Newton Methods provide gains in reliability as long as the stiffness matrix is not singular or ill-conditioned. But it does not help in the vicinity of critical points. For example if $\mathbf{K}$ is exactly singular, system (21.10) is not solvable by the double RHS method discussed in the previous Chapter, and the variable steplength device does not help.

Critical points may come in many flavors. In order of increasing traversal difficulty: isolated limit points, isolated bifurcation points, initially singular structures, and clustered limit and/or bifurcation points.

For the less difficult cases, moving away slightly from the singularity often works. Much tougher is the case when stiffness matrix at the start of the analysis, or of an analysis stage, may be highly singular. This happens, for instance, in some cable, pneumatic and biological structures that are mechanisms in the reference configuration and acquire stiffness as they deform. For such cases a variants collectively known as the Damped Newton Method or DNM, can be effective at the coist of programming complexity. [The name of Regularized Newton is also applied.]

DNM overcomes the singularity problem by adding a diagonal correction to the stiffness matrix. Instead of (21.9) one solves

$$\begin{bmatrix} \mathbf{K}+\gamma\mathbf{D} & -\mathbf{q} \\ \mathbf{a} & g \end{bmatrix} \begin{bmatrix} \mathbf{d} \\ \eta \end{bmatrix} = \begin{bmatrix} \mathbf{r} \\ c \end{bmatrix}, \tag{21.12}$$

where $\mathbf{D}$ is a nonnegative diagonal matrix $\gamma \geq 0$ is a “numeric damping” coefficient, and k supercripts have been omitted. The correction can then be applied with a steplength h:

$$\mathbf{u}^{k+1} = \mathbf{u}^k + h\mathbf{d}, \qquad \lambda^{k+1} = \lambda^k + h\eta. \tag{21.13}$$

If the damping coefficient γ is zero and h is unity, the CNM results. As γ is increased the method approaches steepest descent if $\mathbf{D} = \mathbf{I}$ and scaled steepest descent for general $\mathbf{D}$. This has been useful in conjunction with cable net structures that must traverse highly singular regions. Two choices for $\mathbf{D}$ tried in that case are

$$\begin{aligned} \mathbf{D} &= \beta\mathbf{I}, \qquad \beta = \frac{\mathbf{r}^T\mathbf{K}\mathbf{r}}{\mathbf{r}^T\mathbf{r}} \\ \mathbf{D} &= \beta\mathbf{D}_K, \qquad \beta = \frac{\mathbf{r}_K^T\mathbf{K}\mathbf{r}_K}{\mathbf{r}_K^T\mathbf{r}} \end{aligned} \tag{21.14}$$

where $\mathbf{D}_K = \mathbf{diag}\,\mathbf{K}$ and $\mathbf{r}_K = \mathbf{D}_K^{-1}\mathbf{r}$. Practical values for the damping coefficient γ may be characterized as follows:

$\gamma \geq 1$	very heavy damping
$1 \geq \gamma \geq 0.1$	heavy damping
$0.1\gamma \geq 0.01$	moderate damping
$0.01\gamma \geq 0.001$	light damping

The best results for cable net structures were obtained with light damping. Once the structure acquires sufficient stiffness by deforming, the correction terms may be removed by setting $\gamma = 0$.

§21.5. Chord and Modified Newton Methods

The problem of high computational cost of CNR per step can be alleviated if the same stiffness matrix is maintained for several iteration steps. This general class of methods, collectively known as *chord methods* is based on the iteration scheme

$$\begin{bmatrix} \overline{\mathbf{K}} & -\bar{\mathbf{q}} \\ \mathbf{a}^T & g \end{bmatrix} \begin{bmatrix} \mathbf{d} \\ \eta \end{bmatrix} = - \begin{bmatrix} \mathbf{r} \\ c \end{bmatrix}, \tag{21.15}$$

$$\mathbf{u}^{k+1} = \mathbf{u}^k + \mathbf{d}, \qquad \lambda^{k+1} = \lambda^k + \eta \tag{21.16}$$

Here $\overline{\mathbf{K}}$ and $\bar{\mathbf{q}}$ denote an approximation to $\mathbf{K}$ and $\mathbf{q}$ in some sense, which is maintained fixed for several or all iteration steps. On the other hand $\mathbf{r}$ and c are changed at each iteration. Several variants result of this general scheme result according to two criteria: (a) How $\overline{\mathbf{K}}$ and $\bar{\mathbf{q}}$ are chosen and updated. (b) How $\mathbf{a}$ and g are chosen and updated.

Two specializations of the chord method have proven effective in practice. If $\overline{\mathbf{K}} = \mathbf{K}_n$, which is the stiffness matrix at the start of the n^{th} increment, which is kept fixed thereafter, the *modified Newton method* (MNM) method results. There is a variant called *delayed modified Newton method* (DMNM) for which $\overline{\mathbf{K}} = \mathbf{K}^0$, which is the stiffness matrix evaluated *after* the predictor step, and which again is kept fixed for all k.

Updating versions of MNM and DMNM, identified by acronyms UMNM and UDMNM, respectively, emerge if $\overline{\mathbf{K}}$ is allowed to vary during the iterative process. Several strategies to that effect can be devised. Only three, ranging from the simplest to the most sophisticated, are mentioned here:

1. *Periodic update*: Recompute $\overline{\mathbf{K}}$ every $m \geq 1$ iterations. The “period” m is chosen on the basis of prior experience, relative computational cost of factorization versus solving, etc. Obviously $m = 1$ gives back the conventional Newton method.

2. *Residual monitoring*: If the residual norm $||\mathbf{r}^k||$ does not steadily decreases over a certain “subperiod” m^*, $\overline{\mathbf{K}}$ is recomputed. Typically $m^* = 3$ or 4, which allows for “residual spikes” common as $\overline{\mathbf{K}}$ is reset. This strategy is best combined with the previous one by choosing m as a multiple of m^*.

3. *Progressive update*: This merges chord methods with a nonunitary steplength h.

§21.6. Quasi-Newton Methods

Quasi-Newton (QN) methods represent a refinement of the MNM methods. The stiffness $\overline{\mathbf{K}}$ is updated at each iteration step with rank-one or rank-two matrices built up from information from the previous iteration. In this way a better approximation of the actual stiffness matrix is obtained while still avoiding revaluation and factorization.

The idea comes from the field of optimization, where QN methods (also called *variable metric methods*) have enjoyed great success. They were proposed for solving nonlinear structural problems in the late 1970 with high hopes. Evidence shows, however, that the moderate improvements in reducing the number of iterations to convergence does not compensate for the increase in programming complexity and storage. The idea has some uses, however, in the derivation of accelerator and secant formulas presented in Chapter 22.

§21.7. *Convergence of Modified Newton

(ASEN 5107 students pls ignore this advanced material. It is placed here for eventual development)

Assuming for simplicity that λ is kept fixed, then the limit relaxation equation becomes

$$\overline{\mathbf{K}}\dot{\mathbf{u}} = -\mathbf{K}(\mathbf{x}) \tag{21.17}$$

where $\mathbf{x} = \mathbf{u} - \mathbf{u}(\infty)$ is the distance to the equilibrium solution at $t = \infty$. This can be modally decomposed as

$$\dot{y} = -\mu y \tag{21.18}$$

in which μ are the roots of the symmetric eigenproblem $\mathbf{K}_0\mathbf{z} = \mathbf{K}\mathbf{z}$ and y are modal amplitudes. The appropriate eigenvalue for the Newton direction $\dot{\mathbf{u}}$ can be estimated by the Rayleigh quotient

$$\mu = \frac{\dot{\mathbf{u}}^T\mathbf{K}\dot{\mathbf{u}}}{\dot{\mathbf{u}}^T\overline{\mathbf{K}}\dot{\mathbf{u}}} \tag{21.19}$$

The structural behavior can be characterized as follows.

1. If $\mu < 1$, the structure is *softening* in the mode y;
2. If $\mu > 1$, the structure is *hardening* in the mode y.

For the MNM to converge,

$$|\kappa| = |1 - h\mu| < 1 \tag{21.20}$$

If the structure softens, MNM converges but the converges rate deteriorates unless h is increased (overrelaxation). If the structure hardens, MNM diverges unless h is cut (underelaxation). But the structure hardens in some modes while softening in others, MNM cannot be continued, and a refactoring of the stiffness matrix is called for.

The preceding observations are well know to experienced investigators. They have observed that MNM works quite well in problems when the structure experiences overall softening.

Remark 21.1. To reduce the variation of μ, one may reduce the incremental step, or proceed to reform the stiffness matrix. In some programs the strategy control attempts to cut the incremental steplengths; after two or three unsuccesful attempts the stiffness is reformed.

Homework Exercises for Chapter 21

EXERCISE 21.1 (Very easy, just to get acquanted with the relaxation equation). Verify that (21.2) and (21.3) followed by $\mathbf{W} = \mathbf{I}$ and $h = 1$ lead to the Conventional Newton Method (20.25) for load control.

EXERCISE 21.2 Starting from (21.8) and (21.9) derive the general form of the Relaxed Newton method. Verify that if $\mathbf{W} = \mathbf{I}$ and $h = 1$ this reduces to the Conventional Newton system for general increment control defined by (20.12)-(20.14).

EXERCISE 21.3 Show that the Damped Newton system (21.12) is obtained if one selects $\mathbf{W}^{-1} = \mathbf{I} + \gamma\mathbf{D}\mathbf{K}^{-1}$ in the relaxation equation (21.8) treated by the Forward Euler integrator (21.11).